

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 7, 2002, 23:39:32 ; Search time 5957.55 Seconds
(without alignments)
7094.489 Million cell updates/sec

Title: US-09-911-667A-1
Perfect score: 2562
Sequence: 1 cccacgcgtccggctttgt.....tgaggactgagngccaagtg 2562

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_om.*
- 20: em_or.*
- 21: em_ov.*
- 22: em_pat.*
- 23: em_ph.*
- 24: em_pl.*
- 25: em_ro.*
- 26: em_sts.*
- 27: em_sy.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htgo_hum.*
- 31: em_htgo_inv.*
- 32: em_htgo_rod.*
- 33: em_htg_hum.*
- 34: em_htg_inv.*
- 35: em_htg_rod.*
- 36: em_htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1459.2	57.0	2421	10	AF060173	AF060173 Rattus no
2	1118	43.6	2104	9	HSN802697	AL359592 Homo sapi
3	472.6	18.4	212183	2	AC087893	AC087893 Homo sapi
C 4	211.4	8.3	183481	2	AC011596	AC011596 Homo sapi
C 5	91	3.6	11856	1	AE007266	AE007266 Sinorhizo
6	89.4	3.5	148432	3	AC004642	AC004642 Drosophil
7	89.4	3.5	157851	2	AC020509	AC020509 Drosophil
C 8	89.4	3.5	300542	3	AE003462	AE003462 Drosophil
9	89	3.5	198677	1	AE001863	AE001863 Deinococc
C 10	84	3.3	19834	3	CEK01F9	222175 Caenorhabdi
11	82.8	3.2	10344	1	AE004767	AE004767 Pseudomon
C 12	81.8	3.2	62204	2	AC067757	AC067757 Homo sapi
C 13	77	3.0	11286	1	AE008141	AE008141 Agrobacte
14	74.2	2.9	81020	8	AB026645	AB026645 Arabidops
15	74.2	2.9	81020	24	AC024128	AC024128 Arabidops
16	70	2.7	13462	1	AF218267	AF218267 Pseudomon
17	69.6	2.7	2622	10	AF060174	AF060174 Rattus no
18	69.2	2.7	1638	9	AF052188	AF052188 Homo sapi
19	69.2	2.7	4948	9	AB018278	AB018278 Homo sapi
20	66.2	2.6	2315	5	DYGSV2TRAN	L23403 Discopysge o
21	66.2	2.6	3190	9	BC000776	BC000776 Homo sapi
22	66.2	2.6	4353	9	AB018279	AB018279 Homo sapi
23	66	2.6	3660	10	RATSV2BA	L10362 Rattus norv
24	64	2.5	35000	1	AB000617	AB000617 Bacillus
25	64	2.5	22160	1	BSUB0002	Z99105 Bacillus su
26	63	2.5	2268	6	AX079057	AX079057 Sequence
27	62.4	2.4	10029	1	AB007908	AB007908 Agrobacte
28	61.6	2.4	16775	1	AB000735	AB000735 Nocardi
C 29	61.6	2.4	40699	3	CEZK637	Z11115 Caenorhabdi
30	59.8	2.3	7424	1	AB050935	AB050935 Pseudomon
C 31	59.2	2.3	11886	1	AE004638	AE004638 Pseudomon
32	58	2.3	159167	2	DMBR42L18	AL122031 Drosophil
33	57.2	2.2	37221	2	AC012886	AC012886 Drosophil
34	57.2	2.2	126340	2	DMBR18C8	AL122030 Drosophil
C 35	57.2	2.2	299537	3	AE003438	AE003438 Drosophil
36	56.6	2.2	3868	4	S47919	S47919 p87-transpo
37	56.6	2.2	7080	1	AF305325	AF305325 Comamonas
38	56.4	2.2	35654	1	SC7H1	AL021411 Streptomy
39	53.8	2.1	17641	1	AE001804	AE001804 Thermotog
40	53.2	2.1	1885	6	A48158	A48158 Sequence 5
41	51.4	2.0	297	9	H518A12R	Z65016 H.sapiens C
42	51	2.0	134092	2	H0702G05	AL442105 Oryza sat
43	50.8	2.0	3663	10	RATSV2A	L01788 Rattus norv
44	50.8	2.0	3844	10	RATSV2AA	L05435 Rattus norv
45	50.2	2.0	2132	10	RNOCTK	X98334 R.norvegicu

ALIGNMENTS

RESULT 1	AF060173	AF060173	2421 bp	mRNA	ROD	21-NOV-1998
AF060173	LOCUS	Rattus norvegicus SV2 related protein (SVOP) mRNA, complete cds.				
DEFINITION	AF060173					
ACCESSION	AF060173.1	GI:3901267				
VERSION						
KEYWORDS						
SOURCE	Norway rat.					
ORGANISM	Rattus norvegicus					
REFERENCE	1 (bases 1 to 2421)					
AUTHORS	Janz, R., Hofmann, K. and Sudhof, T.C.					
TITLE	SVOP, an evolutionarily conserved synaptic vesicle protein, suggests novel transport functions of synaptic vesicles					
JOURNAL	J. Neurosci. 18 (22), 9269-9281 (1998)					
MEDLINE	99019745					
REFERENCE	2 (bases 1 to 2421)					
AUTHORS	Janz, R., Hofmann, K. and Sudhof, T.C.					

TITLE Direct Submission
 JOURNAL Submitted (16-APR-1998) Center for Basic Neuroscience, Department of Molecular Genetics, HHMI, UT Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, TX 75235, USA

FEATURES
 source Location/Qualifiers
 1..2421
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /tissue_type="brain"
 1..2421
 /gene="SVOP"
 301..1947
 /gene="SVOP"
 /note="synaptic vesicle protein; similar to transport proteins"
 /codon_start=1
 /product="SV2 related protein"
 /protein_id="AAC78627.1"
 /db_xref="GI:3901268"
 /translation="MEEDLQRLPVVKFRRTGESARSEDAAAGEHUUVQIEGVKRVG
 LEAVLDGAAAPKREFANPTDDTFMVEDAVEAIGFRQWKLVSILTGLAMADAMEM
 ILSILAPOLHCERLPSQVALLSVVFIMGMSSTLWGNISDOYGRKTLISVFWT
 LYVGLISAFAPVYSWILVRLGVFCIGVPOSVTLVAFELPMKARAKILLIEFWA
 IGTVEVLLAVFVMSLGRWLLLSAARPLVPALCPWLPESARYDVLSGNQERAI
 TUKRIETENGAPWPLGKLIISQEDRGKRRDLFTPHFRWTLILLNFIWFSNFSYGL
 VLLTTELFOAGDVCSISSRKKAVEAKCSLACEYLSKEDYMDLLWLTSEFFQGVLTW
 VIDRLGRKTMALCFVIFSLCSLLIFICIGRNVLTLLLFIARAFISLGGFOAAVYVTP
 VYPTATRALGLGTCGMARVGLITPFIQVLMLESSVYLTLAVISGCCLLAALASCEL
 PIETKRALOESSHREWGEMVGRGTNGVPRNSGSE"
 BASE COUNT 459 a 658 c 735 g 569 t
 ORIGIN

Query Match 57.0%; Score 1459.2; DB 10; Length 2421;
 Best Local Similarity 83.0%; Pred. No. 0;
 Matches 1825; Conservative 0; Mismatches 343; Indels 30; Gaps 13;

QY 41 ggtccctcccaacagctgctccaggaggaagcggcggtgctgtccagctt 100
 DB 168 GGCTCCCTCCCCAGCAGCTGC-----CCAGCGAGGAAGCCGC---GGAGCTGTCCAGCAT 220
 QY 101 ccgagtgctgaacccgagggctgctatccaccactaccatgtagggccatgagaag 160
 DB 221 CCCGGTGCTCAAAAGCCGAGCGCGGCCCTGCTTC---ACCATGTAAGGTGCATGCCCA- 276
 QY 161 ggtcatctcgtgcgacgagcgacatggaagaggaacttattccagctaaaggagctccg 220
 DB 277 GGCTCGCCCTGGCGCAGCGCGACATGGAGGAGGACCTGTTCAGCTCAGGCGAGTTCGCCG 336
 QY 221 gttgaaattccgtgcacagcgagagtgcaaggtcagagagacgacgagcttcagga 280
 DB 337 GTGGTGAATTCGCCGCACAGGAGAGAGCGCAGCGTCACAGGACGACGCGCTTCGCCGG 396
 QY 281 gacatgaagtccagattgaaggggtccacgctgggcccctagagcgtggaagctggatgat 340
 DB 397 GAACATGATGTTTCAGATGTGAGGGGGTCCGAGTGGGCCGTAGAAAGCTGTGCGAGCTGGATGAT 456
 QY 341 ggggcagctgtgccaaaggagtttgccaatccaccacgatgatactttcatggttggaagat 400
 DB 457 GGAGCACTGTGCCCAAGAGAGTTGCCAATCCCACTGATGACACTTTCATGGTTCGAAGAT 516
 QY 401 gcagtgaaagccattggctttgaaaaatttcagtggaagctgtgtgtctcactggcttg 460
 DB 517 GCGGTGGAAGCCATTGTGGTTTCGGAAGATTCCAGTGGAAAGCTCTGTCTTCACCGGGCTTG 576
 QY 461 gcttggaatgactgatgcattggagatgatcctcagcactcctggccacacagctgcat 520
 DB 577 GCTTGGATGCGGACCGCCATGGAGATGATGATCTGAGCATCTTCGGCCCTCAGCTGCAC 636
 QY 521 tgcagtggaaggtcccaagagtggaaggtggcattgtgcacctcgggtgtgttttaggc 580
 DB 637 TGCAGTGGGCACTCCCGCACTGGCAGGTGGCGCTGTGACTTTCGGTGGTCTTCATTGGT 696

QY 581 atgagtlccagctccacgctctggtgggaataatctcagaccagatagcgaggaacacaggg 640
 DB 697 ATGATGTCCAGTTCTACGCTCTGGGGAACATCTCGATCAGTATGCGAGAAACAGGG 756
 QY 641 ctgaagatcagcgtgctgtggactctgtactatgcatccttagtgcatttgcgccgtg 700
 DB 757 CTCGAAGATCAGTGTGTTCGGACCCCTGTACTACGGGATCCTCAGCGCTTTCGGCGCCAGTG 816
 QY 701 tatagctggatcctggtctccggggccctgggtgggcttcgggatacggagagttcccccag 760
 DB 817 TATAGCTGGATCCTGCTCCGAGGCCCTCGTGGGCTTTGGGATTGGAGGGTGCCTCAG 876
 QY 761 tcggtgacgctgatatcccgagttccttcccatgaaagccagagcgtaaatgtatttgcgtg 820
 DB 877 TCGGTGACCTGTACGCCGAGTTCTCCCATGAAGCCGACAGGCCAAGTGCAATTTGCTG 936
 QY 821 attgaggtattctgggccatcgggacagtggttcgaggtcgtctcctggtgctgtgtcgtgatg 880
 DB 937 ATTGAGGTTTTCGGGCCATCGGACCGGTTCGAGGTGCTTCTGGCTGTGTTGTGTGATG 996
 QY 881 ccagcctgggctgggctgtggtgctcatctctcagctgtcccgcctctctctcttgc 940
 DB 997 CCCAGCCTGGGCTGGCGCTGGCTGCTGCTGCTGCTGCGCGCTCCACTACTTGTCTTTGCT 1056
 QY 941 gtgctgtttctgctgctgaaagtgaaggtatgattgctgtcagggagaaacacagga 1000
 DB 1057 GTTCTGTGTTTCTGGCTGCCAGAGAGTGTAGTAGTACATGTGCTGTGCTGTTGGAACACAGGAA 1116
 QY 1001 aagcgaatgcacccttaagagagatagcaactgaaacgagagctcccatgccgctgggg 1060
 DB 1117 AAGCGATTGCTACCTTAAAGCGGATCGCAACAGAAATGAGAGCCCATGCTCTGGGG 1176
 QY 1061 aaactcatctctccaggaaggaagccaggaacaaatgagggagcctttttcacacccat 1120
 DB 1177 AAGCTCATCATCTCCACAGAGGAAGCCGAGGCAAAATGAGGGACCTTTTTCACACCCAC 1236
 QY 1121 tttagatgaaacttctgctgtggtttatatggttttccaatgacttctcttactac 1180
 DB 1237 TTTCTGTGGACAACCTGCTGCTGTGTTTATATGTTCTCCAATGCCCTTCTCTATTAC 1296
 QY 1181 gggtagtctctaccacagaaactctccagcgggagatgctgcggcatctccag 1240
 DB 1297 GGCTGTGTTCTGCTACACAGAACTCTTCAGCGCGGAGATGTTTCAGCATCTCCAGC 1356
 QY 1241 cgaaagaagctgtagagcgaataatgcagcctggcctgcagtaacctgagtgaggagat 1300
 DB 1357 CGGAAGAAGCGGTGGNAGCCAAATGCAGCTGGCTGTGTGAGTACCTCASCAGAAGAGAT 1416
 QY 1301 tacatggaactgtgtgagaccacctctctctgagtttccagggtgctcttgatctgtg 1360
 DB 1417 TACATGACCTGCTGTGGACACCTGCTGTGAGTTCCAGGTGCTCTGTGACTCTGTGG 1476
 QY 1361 attattgacgcctgggcgcaagaagaccatggcctgtgctttgtcatctctctctc 1420
 DB 1477 GTCATGACCGCTTGGCGCCCAAGAGACCATGGCTCTGTGTTTGTGTCATCTTTTCCTC 1536
 QY 1421 tgcagcctcctgctgtttatctgtgttggaagaaatgtgctcactctgttactcttcatt 1480
 DB 1537 TGCAGCCTCCTGCTGTTCATCTGCATTGGAAAGAAATGTGTAACCCCTCTTTACTGTTCA 1596
 QY 1481 gcaagagcgtttattcttgaggccttcaagcggcatatatttacacacctgaggtctac 1540
 DB 1597 GCAAGACGTTTTATTCTGAGGCTTCCAAAGCAGCCCTACCTTTACACGCTGAGGTGAT 1656
 QY 1541 cccacggaacgcgggccctcgccctgggcacctgcagcggcctggaagagtggtgtgct 1600
 DB 1657 CCAACGGCAGAGGGCGCTGGGCCCTGGGCACCTGCAGCGGCATGGCGAGAGTGGCGCG 1716
 QY 1601 ctcatcactcgttcatcgcgcccggtgtagtctggaatcctctgttacctgactctgga 1660
 DB 1717 CTCATCACTCCATTCATAGTCAAGTCAAGTGAATGCTGGAATTTCCCGTACCTGACCCCTGGCC 1776
 QY 1661 gttacagtggtgctgctcctcctggctgcctcctcctgcttttttgcacattgagacc 1720

QY 2115 aatgccagcaacaaatcagacacacccagagtcacccggccaaacccctcagtgaacaac 2174
 Db 721 AATGCCAGCAACCAATCAGACACACCCAGAGTCACCCGGCCAAACCCCTCAGTGAACAAC 780
 QY 2175 caaaatattctctgtagatcacgtccagctcagcccaatgtgacacatgctgtccaac 2234
 Db 781 CAAAATATCTCTGTAGATACCGCTCAGGCTCAGGCCCATGTGACACCTGCTGTCCACC 840
 QY 2235 caccgagctgttcagtagatttctccacacacacagcccccaggctttcttcttggaaa 2294
 Db 841 CACCGGACCTGTTTCAGTAGTTCCTCCACACCCACAGCCCCAGGCTTCTTCTTCTTGA 900
 QY 2295 ttgagggcatcaggtgtgtgtgagcagctatttctgagggggcccccgggtttgc 2354
 Db 901 TTGAGGGCATCTAGGTGTGTGTGAGCAGCTATTTCCTGGCAGGGGCCCCCGGTTTGC 960
 QY 2355 ctccctagaacctgacagtggaattctgacagatgagacatgtgcatcacaacaggag 2414
 Db 961 CTCCTAGAGCCCTGACAGTGGATCTCTGGCAGATGGATGGTTCATTAACACTGGAG 1020
 QY 2415 ccacatgcccccacccagccctnttggagttgcccgtttgtggcaccagagatccaga 2474
 Db 1021 CCACATGSCATCCACCCAGCCCTCTGGAGTGTG-CCGTTGTTGGCACCAGAGATCCAGA 1079
 QY 2475 tgtgtctggggagcagctggtgtgtgtgacacaggtgacacacacacacacacacacccc 2534
 Db 1080 TGTGTCTCTGGGACAGCTGTGCTTGCACAGGTGACAACTCAGAGCGCGTTACCCCC 1139
 QY 2535 tggggaactgaggaactgaggaactgaggaactg 2562
 Db 1140 T-GGGAAC-GAGGACTGAGGCGCAAGTGTG 1166

RESULT 3

AC087893
 LOCUS
 DEFINITION Homo sapiens chromosome 12q clone RP11-423G4, WORKING DRAFT
 AC087893
 VERSION
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 212183)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B., Homsif.F., Howard,S., Huber,J., Huiy,K.S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseghe,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,

Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Slisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczkyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.
 Unpublished
 2 (bases 1 to 212183)
 Woriy,K.C.
 Direct Submission
 Submitted (06-FEB-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Jun 8, 2001 this sequence version replaced gi:13928634.

COMMENT

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 Project Information
 Center project name: HCHH
 Center clone name: RP11-423G4
 Summary Statistics
 Sequencing vector: Plasmid; M77789
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 201526 bases at least Q40
 Consensus quality: 205059 bases at least Q30
 Consensus quality: 206774 bases at least Q20
 Estimated insert size: 205251; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 5.3x in Q20 bases; sum-of-contigs estimation
 NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbankdraft_data.html).
 NOTE: This is a 'working draft' sequence. It currently consists of 18 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
 1 41137: contig of 41137 bp in length
 * 41138: gap of unknown length
 * 41238: contig of 27796 bp in length
 * 69034: gap of unknown length
 * 69134: contig of 20257 bp in length
 * 89391: gap of unknown length
 * 89491: contig of 10690 bp in length
 * 100181: gap of unknown length
 * 100281: contig of 12507 bp in length
 * 112788: gap of unknown length
 * 112887: contig of 14851 bp in length
 * 127839: gap of unknown length
 * 127838: contig of 9738 bp in length
 * 127839: contig of 9738 bp in length
 * 137577: gap of unknown length
 * 137676: contig of 7370 bp in length
 * 145047: gap of unknown length
 * 145147: contig of 9686 bp in length
 * 154833: gap of unknown length
 * 154933: contig of 9758 bp in length
 * 164691: gap of unknown length
 * 164791: contig of 7422 bp in length
 * 172213: gap of unknown length
 * 172313: contig of 6961 bp in length

VERSION
KEYWORDS
SOURCE
ORGANISM

AC011596.22 GI:15029431
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 183481)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaral, C., Amaratunga, H.C., Arc, J.R., Banks, T., Barbra, J.,
Benton, J., Bimonte, K., Blankenburg, K., Bonnin, D., Bouck, J.,
Bowle, S., Brieva, M., Brown, M., Brown, M., Bryant, N.P., Buhay, C.,
Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carton, T.F.,
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Davy-Carroll, L., Dederich, D.A., Delaney, K.K., Delgado, O.,
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
Garza, N., Gill, K., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, K., Harris, K., Hart, M., Havlak, P., Hawes, A.,
Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C.,
Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
Louliseed, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M.,
Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S.,
Ogih, M., Okunju, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
Peery, J., Peters, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojebokan, I., Rolfe, M.,
Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N.,
Sisson, I., Sodergren, E., Sonaik, T., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Watlington, S., Williams, G., Williamson, A., Wleczka, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

REFERENCE
AUTHORS

FEATURES
source

Location/Qualifiers
1. 212183
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12q"
/clone="RP11-42364"
BASE COUNT 55581 a 50102 c 50442 g 54260 t 1798 others
ORIGIN

Query Match 18.4% Score 472.6; DB 2; Length 212183;
Best Local Similarity 90.7% Pred No 1.2e-112;
Matches 525; Conservative 0; Mismatches 51; Indels 3; Gaps 2;

Qy 1586 gcaagagtggtgtctctcatcactccgttcacgagtgatgctggaatcctctgtg 1645

Db 68423 GAAGGGATGCTTTCACCGACCCGCTGCTCCCTAGGTCATGCTGGAATCCTCTGTG 68482

Qy 1646 tacctgaacttggaattacagtggtctgctcctcctgctgctgctgctcctgctt 1705

Db 68483 TACCTGACTCTGGCAGTGTACAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTT 68542

Qy 1706 ttgcccattgagaccagagggagagctgagagtgatgctgagccacccggagtgagg; 1765

Db 68543 TTGCCATTGAGACCAAGCGGAGGACTGAGGAGTCCACCCACCGGAGTGGGCCAG 68602

Qy 1766 gagatggtcggccgaggaatcacagtgatgctgagtgatgctgagtgatgctcag 1825

Db 68603 GAGATGCTGGCCGAGGAATCACGGTGCAGGTGTACAGGTGCAACTCTGGCTCTCAG 68662

Qy 1826 gaatgagcagatggggagctgagctggtcttggagctgagagctgaggggctggc 1885

Db 68663 GAATGAGCAGATGGGGAGTACGCTGCTTTGAGGCTGCAGAGCTTGGGGGCTGGC 68722

Qy 1886 agggcccaactggggcactgattgctacgtgcagacatcaagaactcaccgaagatga 1945

Db 68723 AGGCCCAACTGGGCACTGATTGTCTACTGCGGACATCAAGAACTCACCAGAGTATGA 68782

Qy 1946 cctgacacacaggtttgtgctgctgactcagttgctcattgctcattgaggtccacc 2005

Db 68783 CTGACCAACAGAGGTTTGTCTGTGACTCACTTGTCTCACTTCACTTCACTTCACT 68842

Qy 2006 agggat-gggagatgttctgctcctagggggttctctgtatagtgtgaaagcttgctc 2064

Db 68843 AGGGATGGGGAGATGTTTGTCTTANGGGGTTCTCTGNTATGGGCTGAAAGCTCTGTC 68902

Qy 2065 ataacctgtggtacatagtggaacactaccatattagaggttctgtatgacagca 2124

Db 68903 ATAACTGCGGATCTACATCGGGAGACACCCATATATAGAGCGGTCTGCGCATCGCATCA 68962

Qy 2125 accaatcagacacccacccagagtcacccggcccaaccct 2163

Db 68963 ACAATTC--AAACCATCAAACTCACCAGGCTCAACCCCT 68999

RESULT 4
AC011596/c
LOCUS
DEFINITION
SEQUENCE, 8 unordered pieces.
AC011596

AC011596 183481 bp DNA
Homo sapiens chromosome 12 clone RP11-117B7, WORKING DRAFT
31-JUL-2001
HTG
AC011596

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Unpublished
2 (bases 1 to 183481)
Worley, K.C.

Direct Submission
Submitted (08-OCT-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 30, 2001 this sequence version replaced gi:14547725.

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Project Information
Center project name: HMKK
Center clone name: RP11-117B7

Summary Statistics
Sequencing vector: Plasmid; M77789
Sequencing vector: M13; L08821

Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 208349 bases at least Q40
Consensus quality: 216913 bases at least Q30
Consensus quality: 216913 bases at least Q20
Estimated insert size: 193550; sum-of-contigs estimation
Quality coverage: 9.2x in Q20 bases; sum-of-contigs estimation


```

similarity to short chain"
/codon_start=1
/transl_table=11
/product="Probable"
/protein_id="AAK65450.1"
/db_xref="GI:14523918"
/translation="MOLKSRVIVTGAASGLGAATVPMLAQEGATVLGLDLKPAGGEE
PAELGAAVFRNADVTNEADATAALAFKQEFQGHVHCLVNCAGTAPCEKILGRSGPH
ALDSFARTVAVNLIGTFNEMIRLAENVMSQGEPAQDGERGVIVNTASTIAAFDQGIQQA
YAASKGVAALTLPAAELARFGIRVVVTIAPGIFDTPMAGMFQDVODALAASVPFEP
RLGRAEYAAALVKHICENTMLNGEIVRLDGLALMAPR"
3995..5065
/gene="Sma1454"
/CDS
3995..5065
/gene="Sma1454"
/function="Miscellaneous; Not classified regulator"
/note="glimmer prediction; global similarity to OruR from
P. aeruginosa (GI: 1657970) which regulates ornithine
utilization; pfam domain similarity to N-terminal
helix-turn-helix motif"
/codon_start=1
/transl_table=11
/product="Putative transcription factor"
/protein_id="AAK65451.1"
/db_xref="GI:14523919"
/translation="MFNEKQPSHWRMTAPCFMDDTLECLRHRHGVYDAGPLLAQAGLP
SIVTGPVSANQYAFHVAHQAQMDDEFFGEGARPMRAGSPALLCHAILSTATLEHALR
RAKRLVLENPHAGELVYENGLAQIVLKDAGATRSFAFYRTFMIILHGVNCLWIGRR
LPTRWDVFCSPAGDYRLFFGAPHFDDQRTRLVFDAYELKLPPIRDERALKHFL
RHAPANILLYRHIDAGLSAAIRQLRHATAPSAWPGFEATAARMRIAAPTLLRRLRQEG
QTVRSIKEDLLRALAMEALADGRNTNAQLAVELCFSEPSAFHRAFRKWTGSKSPAQFR
NATOTDPAEISSPRKRTQPOES"
complement(5047..5565)
/gene="Sma1455"
/CDS
complement(5047..5565)
/gene="Sma1455"
/function="Miscellaneous; Unknown"
/note="glimmer prediction"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK65452.1"
/db_xref="GI:14523920"
/translation="MRFRTGCVYRAHDPWAFKPAASGDGAALKARENPKGVKTYLIAL
SMTAVKEANQGFARHIDPCVLCSDYDIDCADIALDTAERCRAEHGVSLMEACSWAGA
FAECRSPASWALYNRLHPRCTAGILVPSFAPGTGEAGDRNLVLTWPGDLPFLRLDVTDP
SGRLPKDQLSWG"
complement(5562..5951)
/gene="Sma1456"
/CDS
complement(5562..5951)
/gene="Sma1456"
/function="Miscellaneous; Unknown"
/note="glimmer prediction"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK65453.1"
/db_xref="GI:14523921"
/translation="MAQAKIQSDGGPLILSYMDKGGKIAVQOVAVFGMGSKTQLAET
AGLARLTLRPKSRGKTQNRLEMLIISRVTDWAGKEQAMAWYRAQPLPAFGRR
TABALVKDGKAAAVRDYLDHMLGGFA"
complement(6059..7063)
/gene="Sma1459"
/CDS
complement(6059..7063)
/gene="Sma1459"
/function="Small molecule metabolism"
/note="glimmer prediction; global similarity to
malate-dehydrogenase from M. fervidus (GI: 80950)"
/codon_start=1
/transl_table=11
/product="Putative"
/protein_id="AAK65454.1"
/db_xref="GI:14523922"

```

```

/transl_table=11
/product="Putative muconate cycloisomerase"
/protein_id="AAK65455.1"
/db_xref="GI:14523923"
/translation="MKVISNVRVPLVLPKQPHWSYGIRESFAVNIIEIAUDGTV
GIGECTVAPQDTGTATILYRLAKHLVGHSPHDVAPLARIHFOEYILGHGANIMRAANO
IFSGIDMAMMDLQKLAGLPVHOLLGGAHRKAVGYFFLQGETAEELARDAAVGHAOG
ERYFYLVKVGREGKLDLEITAAVRGEIGDARLRLDANEGWSHIDAINCRKLEKYDIEF
IEQPTVSWSPANAHVREKVGIPTVAIQAAFTLYDVYIEICRRAADMICIGPREIGGI
OPMKAAVAEAAAGAKTICIHSSFTTITTCAEHIIIGLIAINLDDGNGIMQOLVQEDIV
SSPDLTPKNGWLDAPFKPGIGLQALAEPLVAEGEGRYAASR"
complement(8311..9201)
/gene="Sma1462"
/CDS
complement(8311..9201)
/gene="Sma1462"
/function="Cell processes; Transport of small molecules"
/note="glimmer prediction; global similarity to
choline-binding protein precursor from B. subtilis
(GI:7404430) and several putative ABC transporter
proteins; adjacent ORFs encode other putative members of
the choline transport system"
Query Match 3.6%; Score 91; DB 1; Length 11856;
Best Local Similarity 48.3%; Pred No. 8.9e-13;
Matches 324; Conservative 0; Mismatches 335; Indels 12; Gaps 2;

Qy 383 acttctatggtggaagatgcagtggaagccattggctgttgaaaaatttcadgtgaaagctg 442
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1423 ACAATCAGTAGACGATCGCTCGATCGCGCGGACGGGAACCTTACCAGCGCAGACTG 1364

Qy 443 tctgttctcactggttggcttgatggtgcacatgagatgatgcctcagcatc 502
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1363 ATGCCCATATTCGGTCTGGTGGCTGCCGATGCCATGCCAGGTGCTGGCGTTCGCTTC 1304

Qy 503 ctggcaccacagctgcattgcagtgaggtctccacagctggcaggtggcattgctgacc 562
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1303 ACCGCCCATCGATCGCGCGCCACTTTTCGGCTGACATGTCGCGAGGCACTTCACACGGGC 1244

Qy 563 tcggtggtctttagcatgatgtccagctccacgctctgggaaaaatatctcagaccag 622
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1243 ACGCTTTCTTCTCGGAATGCTGTTCCGGCGCGCGGTTTCGGCAGGCTCGCGGACCGG 1184

Qy 623 tacggcagaaaaacagggctgaagatcagcgtgctgtggactctgtactatggcatcctt 682
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1183 ATCGCGCGCGCGCGCTTCGATCGCCACCGCTGCCATGCCCTTCCTGGGTCTGCTG 1124

Qy 683 agtgcatttgcgcctgtatagctggtctgctccggggcctggtggtcttcggg 742
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1123 TCGCTCTTTCGCGCAGGACTTACCGTTTTCGCTGCTGGGTTTCACGCGGCGCAGCT 1064

Qy 743 atcgaggagg---agttcccccagtcggtgacgctgtatgccagttccttccatgaaagcc 799
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1063 GTCGCGGCGCACACTGCGCGGTTCGACTATCGCATGATGCGGAGTTCTCTCGCGCGCGGAAC 1004

```


Db	107182	GATACCTACACTGTTACGACGCGATCAATGCCTTCGGTTCGGATGTTTCACGTCAG	107241
Qy	440	ctgtctgttctcactggcttggcttggatgctgcatgcatgccatggagatgatgatctctcagc	499
Db	107242	CTGTGCGCTTGTGGGGCTCTGCTGGATGTCGCACTCCCACTCCATGGAAATGGCCATCTCTCAGC	107301
Qy	500	atctcggcaccacagctgcatgagtgaggctcccaagctggcaggggagcattgctg	559
Db	107302	ATATTGGGCCCGCTTTGTCTCGGAGTGGGAATGTCACCAAGTCCACGACGGCATCGCTC	107361
Qy	560	acctcgggtgctttagcgatgatgctccagctcccgctcttggggaataatctcagac	619
Db	107362	ACAACAGTGGTGTCTTTGGGGATGATGTTGAGTCCAGTCTTTTGGCACACAGCTGAGCAAC	107421
Qy	620	cagtcagcgaggaagaaacag	638
Db	107422	AGATATGGCCGGAAGTCGG	107440
RESULT	8		
LOCUS	AE003462/c		
DEFINITION	Drosophila melanogaster genomic scaffold 14200001338603R section 11 of 15, complete sequence.		
ACCESSION	AE003462 AE002575		
VERSION	AE003462.1	GI:7291637	
KEYWORDS	HTG.		
SOURCE	fruit fly.		
ORGANISM	Drosophila melanogaster		
REFERENCE	Adams,M.D., Celiker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F., George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N., Sutton,G.C., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X., Brannon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D., Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Helt,G., Nelson,C.R., Gabor Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J., Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A., Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y., Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D., Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C., Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M., Dodson,K., Doup,I.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferrieria,S., Fleischmann,W., Foster,C., Gabrielian,A.E., Garg,N.S., Gu,Z., Gelbart,W.H., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z., Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J., Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J., Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z., Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C., Kravitz,S., Kuop,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A., Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattet,B., McIntosh,I.C., McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C., Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L., Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K., Nusker,D.R., Pacle,J.M., Palazzolo,M., Pittman,G.S., Pan,S., Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K., Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I., Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C., Stapleton,M., Strong,R., Sun,E., Svirkas,R., Tector,C., Turner,R., Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A., Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T., Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F., Zaveri,I.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H., Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O., Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.		
AUTHORS	The genome sequence of Drosophila melanogaster		

JOURNAL	Science 287 (5461), 2185-2195 (2000)
MEDLINE	20196006
REFERENCE	2 (bases 1 to 300542)
AUTHORS	Adams,M.D., Celnikier,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
TITLE	Direct Submission
JOURNAL	Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA
FEATURES	Location/Qualifiers
source	1..300542
	/organism="Drosophila melanogaster"
	/db_xref="taxon:7227"
	/chromosome="2R"
	/complement(join(8417..8923,14564..14687,19965..>20155))
mRNA	/gene="CG11169"
	/product="CT31208"
	/db_xref="FLYBASE:FBgn0011169"
	/db_xref="FLYBASE:FBgn0034905"
gene	/complement(<8417..>20155)
	/gene="CG11169"
	/map="60A1-60A2"
	/db_xref="FLYBASE:FBgn0011169"
	/db_xref="FLYBASE:FBgn0034905"
CDS	/complement(join(8417..8923,14564..14635))
	/gene="CG11169"
	/note="CG11169 gene product"
	/codon_start=1
	/db_xref="FLYBASE:FBgn0011169"
	/db_xref="FLYBASE:FBgn0034905"
	/protein_id="AAF47060.1"
mRNA	/db_xref="GI:7291638"
	/translation="MSSPLVLTALLVSTICQGLPDLHKQMTPEQLQSVFHVDTIID AVPHYELVQLQHNNHNRNRSSIGKPKVNAALPPHHVKKDLKSNAYISELKHEALA SGGNHLFAPEVSAIKSHNSFSAFQGHNLNLSRATQGLFKCAPIQLRMWTVGSEPNAT HGLDLQEIIVHEQVSRVYSMSIQLKKIYHKRY"
	/complement(<13286..13760)
	/gene="CG15800"
	/product="CT15511"
	/db_xref="FLYBASE:FBgn0015800"
	/db_xref="FLYBASE:FBgn0034904"
gene	/complement(<13286..>13760)
	/gene="CG15800"
	/map="60A1-60A2"
	/db_xref="FLYBASE:FBgn0015800"
	/db_xref="FLYBASE:FBgn0034904"
CDS	/complement(13287..13760)
	/gene="CG15800"
	/note="CG15800 gene product"
	/codon_start=1
	/db_xref="FLYBASE:FBgn0015800"
	/db_xref="FLYBASE:FBgn0034904"
	/protein_id="AAF47061.1"
mRNA	/db_xref="GI:7291639"
	/translation="WVRPETHDGTISVNLALLEKSLVIREMCRGVQVPEDEFLVPL HGISNVLKILLWAEYHKEHEEPWDSKDLPAEELELQISDWDKEFLVEIDSLC KIMEGCNVLDIPWLYKLCIAOKLVLSREPETQSGVVEKIPRFODHFIQLQKE"
	/complement(<26883..27038,27100..>27264))
	/gene="CG13561"
	/product="CT32939"
	/db_xref="FLYBASE:FBgn0013561"
	/db_xref="FLYBASE:FBgn0034906"
gene	/evidence=not_experimental
	/complement(<26883..>27264)
	/gene="CG13561"
	/map="60A1-60A2"
	/db_xref="FLYBASE:FBgn0013561"
	/db_xref="FLYBASE:FBgn0034906"
	/evidence=not_experimental
	/complement(join(26883..27038,27100..27264))
CDS	/gene="CG13561"
	/note="CG13561 gene product"
	/codon_start=1
	/db_xref="FLYBASE:FBgn0013561"
	/db_xref="FLYBASE:FBgn0034906"

/codon_start=1
 /transl_table=11
 /product="oligodipeptidase F, putative"
 /protein_id="AAF12423.1"
 /db_xref="GI:6460718"
 /translation="MLARCLPLRGAGPRSCRPAPSTQSPGCMERERWRALPPMTTQ
 SNLEAVEHKLWVPADDAQWASVAFPAALQEAIDLTAGDVPAMLERWNSLSARLOGSGS
 KLSTHADLHTDDALQQRQTFVFAEVLPQAEASQALTEKLLAVPGYVGPDPFALNYR
 RFQDAALFRANVDLGNHNOOINRHSVTGNQKVLGGQELTLPQAKQOMDSDFRA
 QREARWALSAPNAIAPELDAVMULDILTRQLAHNADEANFRYRKLEDRVDYTP
 EQCLDFHASVDEYVLLSGVGLDITARGLSESLRPWDYNNRNLDPGECRESLRPFAT
 GADLEALAQOATESDPLDAARFQMRGSLDLESRPCKMTHAYCNYPTNEPFLV
 MNVGTAEDEVRLVFEVGHAFHGFYSGDROPLVWNRWSPIEFEVIPSMAWELTLDHL
 SHVFSDDELARYRKQLEGVIAFLPRAQMDAFQHLVAAEPEDVTIADLDAKWLELD
 RTFFPIOWSGLDSEVRKAGWHYHIFQVPFYIYEAICYLAAVAIWRAARDQPAAL
 GRYKDALRLGNTVSPPELYRAAGAEFRFREDYIRLGMFLKEQLG"
 complement(2005..3645)
 /gene="DRA0207"
 CDS
 complement(2005..3645)
 /note="similar to PID:1653945 percent identity: 49.28;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAF12422.1"
 /db_xref="GI:6460717"
 /translation="MNFVFLSPSRSSVAHALNRASLKEEMSKSRSLSALLGALL
 SACGSPENGQAPQPLPNTPEAEPTSPGKVRGLWVDFAGFGIKTPEEVSTLVDTR
 ANNVTLYVQVGRGDCYCNNAAMPTDDPEVPAGFDPLADVIKAHAKDIEVNAWII
 TIALNNSVTTPKSAEHAFTHGPNAGRDNMILTAKDGTVKAGADMVLDPGHDAAY
 INKMYSAKNTAVGIDMFDRVRYPDYNPVGVPQVOMYNETALARYRAETGATGTPD
 ADPWSANRQCVTLNLRVETALAVKAVPGVAIAAATITYGEGPADEAAFAASRPYAE
 YGDWMLTWKSGVLDQNVMMNTKDFVPLQAGKFDQWNTFAAGLRQRYPHTEQVSGAA
 IYLDNIASVTSQIRKQTQAGLGHAGYSYTPBKDNCKEGRGAEEVLADLKLRLTDAGG
 PPFTVHAPLSPVGLUSGEALRAAGHHSVELYVRAGGTWNDVVAATFSDQGGRYG
 FLNLPQGFADVRVDGKILLPNVQATPGRVTMLEPITLN"
 complement(3751..4860)
 /gene="DRA0208"
 CDS
 complement(3751..4860)
 /note="similar to PID:1001269 percent
 identity: 66.97; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="peptide ABC transporter, permease protein"
 /protein_id="AAF12421.1"
 /db_xref="GI:6460716"
 /translation="MTGGPETRTPLQMAWRRYKSRPGVIAGWLIALVIALLSPLF
 APYNTTQHEGYEQPTVHLIHGRLHRPFGVYSTTKRDPVTFAKKTVPDFAHPLP
 AVFTTVKGEYSFLGQSDRLHFCVQDQDEYKEKVPYFLFGLDGLDFSRLLV
 AOVMTGVVWVLLISFALGILGGVSGYFGWVDTLQIRVWVLLSFPRPLILALAT
 LIPARWSPVWVTVGIVAVIALIGWASLVRQVLAARNLDYVSAOAMGASNLKVI
 LHPIPNLSFLILVTATLPGYILGESLTSFLGLGKEPMTSWGLLLRDASKLEVVS
 SPWLLWPGFLFVLSVLAIFNPGDALDAADTQSR"
 complement(4857..5843)
 /gene="DRA0209"
 CDS
 complement(4857..5843)
 /note="similar to PID:1001492 PID:1001545 percent
 identity: 62.93; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="peptide ABC transporter, permease protein"
 /protein_id="AAF12420.1"
 /db_xref="GI:6460715"
 /translation="MLCCMLNLYTKRRILGMIPTLVLSVLCFLVLIKLPQSFIDOYL
 EDPRSKEVANITRQGLDQPAWKYLNWITGVYTRDFGYSFASGAPVTSIIGRLV
 GWTFVTAUTLALVPLVPLGYTAFNRGYPQAQIANFLGYVLAIPDFLAALLLV
 LVYRLGGTVNGGLFQFDMIDAPWSFAKVIIDLLAHMIPVLAVGLGVGLMQRMAST
 LDVLSQDYIRAKAGPQGRVWIKHAVRNAINPLISLAGLSLPTISGTTIIISIVS


```

FANLGTAAQSVAAAVVVIDAHASPLMVAMIQVAAALPLVLLLSILLSGVIAIDNHRKKIM
LWGLSFENAMFATLLAFLGYLDPLVLIISLWISLGGSVTIPAWQAANVOQVPAFM
VSDVLNVNVTARAAGPALGGLLSAVGPVAVFLNSFCYNALIAIWWRNDVP
KRSLLPEGILSGVTAALRFTOYSTVTRLVMSRSPAGLSASAVWALLPLLAHRNDGD
AAIYGYMLGALGAILGSTOVSRLRQRIGSSRLISLAGFTLAILLLTLGLVDNLWL
FVLLGGCGWIGALATYNSAVQILVPMWIKARALALYQALYGLAUGSLFWGLHAE
TMTVHGALLAACLLLSAILLYNSRLPMDAASISRAPSMPOPSVFNTRRMQVL
VSIERYTPAERTDFRAAQPLRLRLRNGAERWSLFRDVSNEPMVQBELFLVDNWIQH
LRMLDRMTLADKIVIDNVNTALHAGDPPQIRHCVSYEASSYDTPLVKLSATPPANDEAG
ATAGS"
gene complement(7338..8678)
/ gene="PA3466"
CDS complement(7338..8678)
/ gene="PA3466"
/ codon_start=1
/ transl_table=11
/ product="probable ATP-dependent RNA helicase"
/ protein_id="AAG06854.1"
/ db_xref="GI:9949609"
/ translation="MPRSPLFSQPLHERLLKALLESFSFSEPTVPVAAAIPKALEGHD
LRVTAQTSQKTAAPLPLHLLAEQKPSRLARLILLLPTRELAQQTLEKVERPAQF
TFIKACLTIGGEDEFKVGARLKNPETIITGPRLLLEQNRAGNIPLODIEVLVLDEAD
RMLDMGFADVALANACPAERQTLFSAHSGAGLNKVI AEVLREPOVLNNOVGEL
NENVRQOYITDVAHKEOLLQWLLSNETYTKAIVETNTRVSADRLTGLRANQHKVF
VLHGEKDQKRLAERLKGAVKILVATDVAARGLDVEGLDVLINFDMPRSGDEYVH
LHGKTRAGACGLAISLICHGDMNLMSSVERYLKQNFERRNIKELKAAQYQPKLKAS
GKAAGSKKQDQKRAAKPAKRPAKPSAVVSALGMAPLKRKKPTAE"
8864..10231
/ gene="PA3467"
CDS 8864..10231
/ gene="PA3467"
/ codon_start=1
/ transl_table=11
/ product="probable MFS transporter"
/ protein_id="AAG06855.1"
/ db_xref="GI:9949610"
/ translation="MPDAPASAAERLERPLSPYHRLVFVILIAAFFSDMLAMMTF
LLGSKAFEGDLSQAAGLLASSFFGVIGALSMLADRFGRKPVFOASITVLMLGAS
VLCSTAGDLSLTYSYVLLGIGMGMEPIAOSLSEMI PASRGGKYIALMDGFNPLGF
DVAAGCLSYFLPLGWRKISFVLVLAIPAVFLATRLPIESPRLWEQAGREOARDVLR
LDEARVMSLGLTELPPLRQPRERSRPGFSAFELWSPAYRRRTLTVMGLMFFAL
LGFLYGLTSMWLSAQGGFAVTSQYVYVTLISLAGIPGLCAAMLVESWGRKPSCLVM
LLGGGAMAYAGTQAVFGGSLIILFGLAMQFFLFGMWAVLYTTPLELYPTSRATG
SGFASAVGRIGSLPLVTLVPLVTGOGGVFTLGLCFGVAALVWVAFGIEETRGL
EELAG"
BASE COUNT 1495 a 3698 c 3395 g 1756 t
ORIGIN

Query Match 3.2%; Score 82.8; DB 1; Length 10344;
Best Local Similarity 47.6%; Pred. No. 1.3e-10;
Matches 277; Conservative 0; Mismatches 302; Indels 3; Gaps 1;

QY 427 atttcagtggaagctgtcttctcaactggttgcttggttgatgctgcatgagatg 486
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8920 ATATCACCAGCTGCTTCTGTGATCATCGCGCTGGCGTCTTCTTCGACTGATGACCT 8979
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 487 gatgatcctcagcatctcctggaccacagctgcatgctgagtgagctgcccaagctgcca 546
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8980 GCGGATGATGACTTCTGCTGCTGCATCAAGCCGAGTTCGGGCTCGACTCGCGCA 9039
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 547 gttggtcattgctgacctcgggtgtgtttttagggcagtgatgctcagctcagcactc 9499 606
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9040 GGCCCGGCTGCTGGCGAGCTGAGCTTCTTCGGCATGTTGATPGCGCGGCCCTGTCGGG 9099
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 607 aaatctctcagacagtcagcagaaacacagagctgaagatcacgctgctggaactct 666
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9100 CATGCTCGCGACCGCTTCGGCGCAAGCCGGTATTCAGGCCACATCGTCTCTGGGG 9159
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 667 gtactatggcattccttagtgatcctgccccggtgtatagctggatcctggtgctccggg 726
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9160 COTGGCCAGCTACTCTGTCTCCACCGCGCGGACCTCGACAGCCCTGACCTTCTACCGGCT 9219
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```

QY 727 cctgtgtgggtcctcggtatcgg---aggaggttcccccaagtcggtgaacgtctgtatgccagtt 783
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9220 ATTCTGGGGATCGGCATGGGATGGAGTTTCCCATCGCCAGTCGCTGCTGTCGGAGAT 9279
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 784 ccttcccatgaagccagagactaaatgtattttgtgattgaggtattctgggcccactcg 843
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9280 GATCCCGGCCACGCCCGCGGGAAGTACATTCGGCTGATGGACGGCTTCTGGCCGCTGGG 9339
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 844 gacagtgttcgaggtctgctcgtgctgctgctgctgctgctgctgctgctgctgctgctg 903
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9340 CTTCTGCTCGCTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9399
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 904 gctcatcctcctcagctgtccccctcctcctcctcctcctcctcctcctcctcctcctcctga 963
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9400 TTTCTGTGCTGCTGCGGCTGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9459
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 964 aaggtcaaggtatgatgtgctgctcaggggaacacaggaagaaagc 1005
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9460 GTTCGGCGCTGCTGTCGACGACGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9501
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
AC067757/c
LOCUS Homo sapiens chromosome 2 clone RP11-423G3 map 2, LOW-PASS SEQUENCE
DEFINITION SAMPLING.
ACCESSION AC067757
VERSION AC067757.1 GI:7651806
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 62204)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 2, clone RP11-423G3
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 62204)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Bennett,L., Bonkhoville,B., Brown,A., Burkett,G.,
Campolongo,A., Castle,A., Choepe,Y., Colangelo,M., Collins,S.,
Collins,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagap,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
TITLE
JOURNAL
COMMENT Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L9993
Center clone name: 423_G_3
```

* NOTE: This record contains 78 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 730: contig of 730 bp in length
* 731 830: gap of 100 bp
* 831 1522: contig of 692 bp in length
* 1523 1622: gap of 100 bp
* 1623 2322: contig of 700 bp in length
* 2323 2422: gap of 100 bp
* 2423 3119: contig of 697 bp in length
* 3120 3219: gap of 100 bp
* 3220 3925: contig of 706 bp in length
* 3926 4025: gap of 100 bp
* 4026 4710: contig of 685 bp in length
* 4711 4810: gap of 100 bp
* 4811 5499: contig of 689 bp in length
* 5500 5599: gap of 100 bp
* 5600 6293: contig of 694 bp in length
* 6294 6393: gap of 100 bp
* 6394 7102: contig of 709 bp in length
* 7103 7202: gap of 100 bp
* 7203 7920: contig of 718 bp in length
* 7921 8020: gap of 100 bp
* 8021 8722: contig of 702 bp in length
* 8723 8822: gap of 100 bp
* 8823 9516: contig of 694 bp in length
* 9517 9616: gap of 100 bp
* 9617 10317: contig of 701 bp in length
* 10318 10417: gap of 100 bp
* 10418 11127: contig of 710 bp in length
* 11128 11227: gap of 100 bp
* 11228 11922: contig of 695 bp in length
* 11923 12022: gap of 100 bp
* 12023 12715: contig of 693 bp in length
* 12716 12815: gap of 100 bp
* 12816 13518: contig of 703 bp in length
* 13519 13618: gap of 100 bp
* 13619 14318: contig of 700 bp in length
* 14319 14418: gap of 100 bp
* 14419 15123: contig of 705 bp in length
* 15124 15223: gap of 100 bp
* 15224 15919: contig of 696 bp in length
* 15920 16019: gap of 100 bp
* 16020 16721: contig of 702 bp in length
* 16722 16821: gap of 100 bp
* 16822 17533: contig of 712 bp in length
* 17534 17633: gap of 100 bp
* 17634 18322: contig of 689 bp in length
* 18323 18422: gap of 100 bp
* 18423 19120: contig of 698 bp in length
* 19121 19220: gap of 100 bp
* 19221 19932: contig of 712 bp in length
* 19933 20032: gap of 100 bp
* 20033 20735: contig of 703 bp in length
* 20736 20835: gap of 100 bp
* 20836 21544: contig of 709 bp in length
* 21545 21644: gap of 100 bp
* 21645 22359: contig of 715 bp in length
* 22360 22459: gap of 100 bp
* 22460 23153: contig of 694 bp in length
* 23154 23253: gap of 100 bp
* 23254 23941: contig of 688 bp in length
* 23942 24041: gap of 100 bp
* 24042 24741: contig of 700 bp in length

* 24742 24841: gap of 100 bp
* 24842 25541: contig of 700 bp in length
* 25542 25641: gap of 100 bp
* 25642 26348: contig of 707 bp in length
* 26349 26448: gap of 100 bp
* 26449 27069: contig of 621 bp in length
* 27070 27169: gap of 100 bp
* 27170 27873: contig of 704 bp in length
* 27874 27973: gap of 100 bp
* 27974 28665: contig of 692 bp in length
* 28666 28765: gap of 100 bp
* 28766 29463: contig of 698 bp in length
* 29464 29563: gap of 100 bp
* 29564 30231: contig of 668 bp in length
* 30232 30331: gap of 100 bp
* 30332 31031: contig of 700 bp in length
* 31032 31131: gap of 100 bp
* 31132 31820: contig of 689 bp in length
* 31821 31920: gap of 100 bp
* 31921 32597: contig of 677 bp in length
* 32598 32697: gap of 100 bp
* 32698 33328: contig of 631 bp in length
* 33329 33428: gap of 100 bp
* 33429 34132: contig of 704 bp in length
* 34133 34232: gap of 100 bp
* 34233 34936: contig of 704 bp in length
* 34937 35036: gap of 100 bp
* 35037 35743: contig of 707 bp in length
* 35744 35843: gap of 100 bp
* 35844 36557: contig of 714 bp in length
* 36558 36657: gap of 100 bp
* 36658 37362: contig of 705 bp in length
* 37363 37462: gap of 100 bp
* 37463 38172: contig of 710 bp in length
* 38173 38272: gap of 100 bp
* 38273 38978: contig of 706 bp in length
* 38979 39078: gap of 100 bp
* 39079 39782: contig of 704 bp in length
* 39783 39882: gap of 100 bp
* 39883 40577: contig of 695 bp in length
* 40578 40677: gap of 100 bp
* 40678 41372: contig of 695 bp in length
* 41373 41472: gap of 100 bp
* 41473 42185: contig of 713 bp in length
* 42186 42285: gap of 100 bp
* 42286 43021: contig of 736 bp in length
* 43022 43121: gap of 100 bp
* 43122 43824: contig of 703 bp in length
* 43825 43924: gap of 100 bp
* 43925 44631: contig of 707 bp in length
* 44632 44731: gap of 100 bp
* 44732 45447: contig of 716 bp in length
* 45448 45547: gap of 100 bp
* 45548 46258: contig of 711 bp in length
* 46259 46358: gap of 100 bp
* 46359 47062: contig of 704 bp in length
* 47063 47162: gap of 100 bp
* 47163 47862: contig of 700 bp in length
* 47863 47962: gap of 100 bp
* 47963 48665: contig of 703 bp in length
* 48666 48765: gap of 100 bp
* 48766 49445: contig of 680 bp in length
* 49446 49545: gap of 100 bp
* 49546 50252: contig of 707 bp in length
* 50253 50352: gap of 100 bp
* 50353 51034: contig of 682 bp in length
* 51035 51134: gap of 100 bp
* 51135 51840: contig of 706 bp in length
* 51841 51940: gap of 100 bp
* 51941 52641: contig of 701 bp in length
* 52642 52741: gap of 100 bp
* 52742 53422: contig of 681 bp in length
* 53423 53522: gap of 100 bp


```

/db_xref="SPTREMBL:Q9LDD1"
/note="hypothetical protein"
/gene="MGH6.7"
/protein_id="AAF35941.1"
/translation="MAYEAHHLNLFSSNREMLHPVEASGVVNTQMYRGTVPTFNPTV
ECQSLSPNITNISPDLRVHQSMMKPTIQSDSLSTFNSNNVDFLRPSRSRKKRSREES
VLWPSAYMOIQKNPTDPLMPLGDDLSNNVQHHFDLIRLSNHEVERMEIEERKRTQ
GRVIVEAQGLKDEIKRAQDEINHIKUNLFLFEKYSCLVCENQIRWDAVQNEATVQ
ALRNSLOOVAVERNWEPPTVADDAQSCGNSDEGDEEEERFKLAGAODTKKNCR
VGMSCRCGGEASVLLPCRHMCCLSCVCGSSLNTPICKSPKTAASHVNLS"
complement(13950..14024)
/rpt_family="(TAAAA)n"
join(<18851..19090,19255..19381,19476..19614,19697..19785,
19881..>20002)
/gene="MGH6.4"
join(18851..19090,19255..19381,19476..19614,19697..19785,
19881..20002)
/codon_start=1
/db_xref="SPTREMBL:Q9LDY9"
/note="hypothetical protein"
/gene="MGH6.4"
/protein_id="AAF35942.1"
/translation="MASSTGILTVAGALLAGDFRLPAVSSLLIPRKTTSSLSCLSNRDLSS
PYNCWLRSLRCKILTTSNKRKFAFGAEQDFLSNVEDTDEMFDDLFRNYGKRVFRS
TDVKSPTAEVDDDAEAFVELAKVASDVKAGDIKVLVPLVYVTRFFLIATAFSRP
QIDAGSRMRDLAEKKYGVANGDVKNPWSVLLDEGDVVIIHFLPPQRTFFNLEDFYGN
AMQIELPEDQSQPRN"
join(complement(21254..22367),complement(<20664..21091))
/gene="MGH6.5"
join(complement(21254..22298),complement(20664..21091))
/codon_start=1
/db_xref="SPTREMBL:Q9M7W5"
/note="unknown protein"
/gene="MGH6.5"
/protein_id="AAF35943.1"
/translation="MMAGNGMPTLGRVKVCDLPSEGLPSDSYKLAVTTLQSLSAQYSA
AIQFPASGALLRGLDSARILYFHQSDSPATNNMIHTNDSQEWCKTSGYYADPQSMQ
ESYEVPLGLTPEPSNMEFFPAGLPDIFALLGKAARVILDAIGFYLNLRSCPTEILD
NYPLENCEVSSVLSCVCCYARPSFPAQHSHLSTDEQLILYSDHDOHLDKSLISFVKSD
KAGLIRDMHGQWILVDVLDPQEAUVYVPGALYQATAGYVSPAVVHTDMLNSLOGSIBG
RFSLAFKLMKSMNLSCSENKRAAGSVEAQQFQLPVSDDFMQRSHSNDELFPNQTLOS
FTVQSDGSMKOLKRRKSDSRCKPLPPSRKRLRLEAQRVLRKVEQIEADKRGIKLRFEC
NLKEENHNHYVNSPCANTRIEIGMPHGVFVPHDLNPKAKIQLFETIEYFGWSETHDM
EFTSLTAQGNQHVTCNHHSSNPFPFAET"
join(complement(24815..25204),complement(24582..24731),
complement(24256..24496),complement(23677..24177),
complement(<23184..23578))
/gene="MGH6.6"
join(complement(24815..25204),complement(24582..24731),
complement(24256..24496),complement(23677..24177),
complement(23184..23578))
/codon_start=1
/db_xref="SPTREMBL:Q9LDF8"
/note="unknown protein"
/gene="MGH6.6"
/protein_id="AAF35944.1"
/translation="MPQYSGTGHCEFTAAASYFSWPTSSRLSNAAEERANYFSLNQKEE
DDDEYSPVETPEPQGRATTLLEIMTI RAFHSKMLRCSYSLGTAIGFIRRGVLTIDP
ADIVYSRVKRWQMLSPQLCPLTALEAGGICWDDVVEFSYFGEPDHPTPKQTFITD
IYDHLQSDPFIIGSGSVASOFTCTGLAIVRSOTGRQGVGTNRHVAVNLIDYPSQKM
FHEPLPAGLPGVYLGAVERATSEITDLDLWFGIFACTNPETVFDAGAFIIPFADDYDLR
VITSYKGVGGEIGEVKATELOSPVGSVLGVKQVYVGRSGLTGTLYALALEYNDERGV
CFITDLVVGHNHRSFPDLEGGSSLIWNKGEERKARPIGIWGTGTSRGLRKLKVGEC
ESWTTGVDLRLTLHQDLITTDGLKAAVQERAASTGTSMSSWADSSPPYVNLKKE
KRSPEEKLEASGLPQVQHIDLEERIEYTKGGAPSVHEHQFMFTFGSCQASAMPETARE
LVAGFTNGSCDGLCVGLRGDGGAKRRRTQVTNERMPAE"
complement(26470..26521)
/rpt_family="(GA)n"
<26932..>27192
/gene="MGH6.7"
26932..27192
/codon_start=1
/db_xref="SPTREMBL:Q9LE44"

```

GenCore version 4.5
Copyright (C) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 7, 2002, 23:46:32 ; Search time 351.36 Seconds
(Without alignments)
6251.330 Million cell updates/sec

Title: US-09-911-667A-1
Perfect score: 2562
Sequence: 1 cccagcgctcgggcttgt.....tgaggactgagngccaaagt 2562

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 930621 seqs, 428662619 residues 1861242
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_ll01.*
1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2560	99.9	2562	21	AAZ49686 Human organic cation
2	529.6	20.7	1906	21	AAZ49689 Rat organic cation
3	447.4	17.5	480	21	AAZ42410 Human 5' EST isola
4	359.8	14.0	401	21	AAA45312 Human secreted exp
5	66.2	2.6	3580	21	AAA16697 Human secreted pro
6	66.2	2.6	4366	22	AAH57556 Human brain cell s
7	63	2.5	2268	22	AAF26353 P. putida oxygenas
8	49	1.9	1356	22	AAH68313 C glutamicum codin
9	49	1.9	309400	22	AAH68534 C glutamicum codin
10	46.6	1.8	1662	19	AAV49559 Human liver cell c
11	46.6	1.8	1888	19	AAV49558 Human liver cell c

12	46.4	1.8	549	21	AAZ49686	Cat flea hindgut a
13	46.4	1.8	1590	21	AAZ49686	Cat flea HMT synap
c 14	46.4	1.8	1590	21	AAZ49686	Cat flea HMT synap
15	46.4	1.8	1875	21	AAZ49686	Cat flea HMT synap
c 16	46.4	1.8	1875	21	AAZ49686	Cat flea HMT synap
17	44.8	1.7	1519	21	AAZ49686	Human OREF ORF2047
18	43	1.7	2191	21	AAZ49686	Rat cerebellar organ
19	41	1.6	1831	20	AAZ49686	DNA encoding a pro
20	41	1.6	1831	21	AAZ49686	Human OCTN2 cDNA s
21	40.2	1.6	758	21	AAZ49686	Trichoderma reesei
22	40	1.5	2135	20	AAZ49686	DNA encoding a pro
23	38.6	1.5	1882	17	AAZ49686	Murine organic ani
24	37.6	1.5	1897	22	AAZ49686	Rat OCT-1 gene. R
25	37.6	1.5	50937	21	AAZ49686	Streptococcus olea
26	37	1.4	470	21	AAZ49686	Human secreted pro
27	37	1.4	6675	20	AAZ49686	Stealth virus nucl
c 28	36.6	1.4	744	16	AAZ49686	Wilson disease gen
29	36.4	1.4	1396	21	AAZ49686	Human OREF ORF1716
30	36.4	1.4	1490	20	AAZ49686	Human organic cati
31	36.4	1.4	1560	21	AAZ49686	Human saccharide-t
32	36.4	1.4	1929	22	AAZ49686	Atherosclerosis-as
33	36.4	1.4	2083	20	AAZ49686	DNA encoding a pro
34	36.4	1.4	2393	22	AAZ49686	Human EST-derived
35	36.4	1.4	2393	22	AAZ49686	Human cDNA encodin
36	36.4	1.4	2402	21	AAZ49686	Human secreted pro
37	36.4	1.4	2460	20	AAZ49686	Human organic cati
c 38	36.2	1.4	710	22	AAZ49686	Human 7TM clone H7
c 39	36.2	1.4	839	20	AAZ49686	Novel nucleotide s
c 40	36.2	1.4	949	22	AAZ49686	Human olfactory re
41	36	1.4	674	21	AAZ49686	Aspergillus oryzae
42	36	1.4	10732	21	AAZ49686	Gene encoding a su
43	35.6	1.4	1960	20	AAZ49686	Soybean hexose car
44	35.6	1.4	3489	21	AAZ49686	Kaposi's sarcoma-a
45	35.6	1.4	3489	22	AAZ49686	Nucleotide sequenc

ALIGNMENTS

RESULT 1
AAZ49686
ID AAZ49686 standard; cDNA: 2562 BP.

XX AAZ49686;

XX 07-APR-2000 (first entry)

DT Human organic cation transporter-like protein (OCTlp) cDNA.

DE Human; organic cation transporter-like protein; OCTlp; transporter;
transmembrane; natriuretic; neuroprotective; neuroleptic; anticonvulsant;
antiParkinsonian; antidepressant; cellular process; cell proliferation;
screen; treatment; prevention; diagnosis; neurodegenerative disorder;
Parkinson's; Huntington; ALS; amyotrophic lateral sclerosis;
CNS disorder; central nervous system; schizophrenia; depression;
behavioural; sleep disorder; eating disorder; Alzheimer's; ss.

OS Homo sapiens.

XX Key Location/Qualifiers
FT CDS 185..1831
FT FT /*tag= a /product= "OCTlp"

XX WO200000633-A1.

XX 06-JAN-2000.

XX 29-JUN-1999; 99WO-US14880.

XX 30-JUN-1998; 98US-0107932.

XX (MILL-) MILLENNIUM PHARM INC.

CC	disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's disease), tumours, bacterial, fungal or viral infections, depression and psoriasis. AAA45926 to AAA45931 represent linker variants which are given in the exemplification of the present invention.
XX	
SQ	Sequence 401 BP; 69 A; 119 C; 104 G; 109 T; 0 other;
	Query Match 14.0%; Score 359.8; DB 21; Length 401;
	Best Local Similarity 98.1%; Pred. No. 5.5e-91;
	Matches 364; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Qy	1386 agaccatggccctgtgcttttgcatttcctctcctctcagcctcctgctgtttatctgtg 1445
Db	31 agaccatggccctgtgcttttgcatttcctctcctctcagcctcctgctgtttatctgtg 90
Qy	1446 ttggaagaatgtctcaactctgttaaccttcatcttcaatgcaagcgattattctgaggct 1505
Db	91 ttggaagaatgtctcaactctgttaaccttcatcttcaatgcaagcgattattctgaggct 150
Qy	1506 tccaagcgcatagtttcacaccttgagtctaccgccggaacgcggccctcgccc 1565
Db	151 tccaagcgcatagtttcacaccttgagtctaccgccggaacgcggccctcgccc 210
Qy	1566 tgggcacctgagcggcatggcaagagtggtgtctcatcactccgttcatcgccacgg 1625
Db	211 tgggcacctgagcggcatggcaagagtggtgtctcatcactccgttcatcgccacgg 270
Qy	1626 tgatgtggaactctgttgtaactgactctggcagtttacagtggtctgctcctctgg 1685
Db	271 tgatgtggaactctgttgtaactgactctggcagtttacagtggtctgctcctctgg 330
Qy	1686 ctgccccctgctcctgcttttgcacattgagaccaagcggaggaactgcaggagtcca 1745
Db	331 ctgccccctgctcctgcttttgcacattgagaccaagcggaggaactgcaggagtcca 390
Qy	1746 gccaccgggag 1756
Db	391 agccactcgag 401
RESULT	5
AAA16697	
ID	AAA16697 standard; cDNA; 3580 BP.
XX	
AC	AAA16697;
XX	
DT	16-JUN-2000 (first entry)
XX	
DE	Human secreted protein clone asl80_1 nucleotide sequence SEQ ID NO:159.
KW	Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW	antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW	antidiabetic; antitachymatic; antiarthritic; antirheumatic; protozoacide;
KW	antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW	infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW	connective tissue disease; multiple sclerosis; erythematosis;
KW	rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW	Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW	insulin dependent diabetes mellitus; graft-versus-host-disease;
KW	autoimmune inflammatory eye disease; allergy; ss.
OS	Homo sapiens.
XX	
PN	WO200009552-A1.
XX	
PD	24-FEB-2000.
XX	
PF	13-AUG-1999; 99WO-US18298.
XX	
PR	14-AUG-1998; 98US-0096622.
PR	17-AUG-1998; 98US-0096815.
PR	04-SEP-1998; 98US-0099229.

PR 23-OCT-1998; 98US-0105368.
 PR 08-JAN-1999; 99US-0115234.
 PR 12-FEB-1999; 99US-0119931.
 PR 18-FEB-1999; 99US-0120575.
 PR 30-APR-1999; 99US-0132020.
 PR 11-AUG-1999; 99US-0096622.
 XX (GEMY) GENETICS INST INC.
 PA Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C.
 PI Metberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
 PI Wong GG, Clark HF, Fechtel K;
 XX
 DR WPI: 2000-205979/18.
 DR P-PSDB: AAY94977.
 XX
 PT New polynucleotides encoding secreted proteins, which may have e.g.
 PT nutritional, chemokine, immune stimulating or suppressing,
 PT hematopoiesis regulating, tissue growth, activin/inhibin
 PT antiinflammatory or tumor inhibition activity .
 XX
 PS Claim 168; Page 613-614; 641pp; English.
 XX
 CC AAA16618 to AAA16697 encode the human secreted proteins given in
 CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
 CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
 CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
 CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
 CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
 CC predicted to have biological activities which would make them suitable
 CC for treating, preventing or ameliorating medical conditions in humans
 CC and animals. The polynucleotides can be used as markers for tissues in
 CC which the protein is preferentially expressed, as molecular weight
 CC markers on Southern gels, and as chromosome markers or tags to identify
 CC chromosomes or to map gene positions. The proteins can be used in the
 CC treatment of immune deficiencies and disorders, such as severe combined
 CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
 CC infections. These infections include human immunodeficiency virus (HIV),
 CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
 CC candidiasis. The proteins can be used to treat autoimmune disorders such
 CC as connective tissue disease, multiple sclerosis, systemic lupus
 CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
 CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
 CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
 CC autoimmune inflammatory eye disease. The proteins can also be used to
 CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
 CC probes for the human secreted proteins from the present invention.
 XX
 SQ Sequence 3580 BP; 734 A; 942 C; 1024 G; 880 T; 0 other;

Query Match 2.6%; Score 66.2; DB 21; Length 3580;
 Best Local Similarity 47.3%; Pred. No. 4.6e-08;
 Matches 234; Conservative 0; Mismatches 258; Indels 3; Gaps 1;
 QY 415 tggcttggaataatcagatgaagctgtcttctcactgcttggcttgatgcttga 474
 DB 504 tggccagccctccagatgagacactgtattgtgcttgctggcgtgatgctga 563
 QY 475 tgcctatggagatgatctcagatcagatcctggcagacagctgctatggagtgagct 534
 DB 564 cgggtggaaggtcttgtggtggtctgctgctgctgagcagctgagacatgtgct 623
 QY 535 cccaagctggcaggtggcattgtgctgctgctgcttcttgtagcagatgctcagctc 594
 DB 624 gtccgactcccaaaagcagatgctaggctcctcctcctcctggcagatgctggagc 683
 QY 595 cagctctggggaataatcagacagctacagcagcaggaagaaacagggctgaagatcgct 654
 DB 684 ctctcctgggaggtctgctgacccgctggctggctggcagctgtgctgctatcgt 743
 QY 655 gctgtggaactgtactatggaactcttagtgcattgctgcccgtgtatgactgact 714

DB 744 ctcaagtcacacagcgtcttctgctctctctctctctctctctctctctctctct 803
 QY 715 ggtgctccggggcctggtggtcttcggatcgcgag--agttcccccagtcggtgagct 771
 DB 804 cttctgcgcctacttctctggtggtggtggtggtggtggtggtggtggtggtggtggt 863
 QY 772 gtagccgagttccttccatgaaagccagacgactaaatgtattttgtgattgaggtatt 831
 DB 864 ttctccgagttctggtcccaaggagaaacagagggagcattgtgagctggtctgcatgt 923
 QY 832 ctggggccatcgggacagctgttcgaggtcgtcctgagctgtgctgagcgcacacgtggg 891
 DB 924 ttgtagtattggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 983
 QY 892 ctggcgttggtgctgct 906
 DB 984 gtggagtttccagat 998
 RESULT 6
 AAH57556
 ID AAH57556 standard: cDNA: 4366 BP.
 XX AC AAH57556;
 XX DT 10-SEP-2001 (first entry)
 XX DE Human brain cell specific cDNA sequence SEQ ID NO:396.
 XX KW Human; tissue specific; diagnosis: brain; heart; skeletal muscle;
 KW lung; liver; uterus; ovary; stomach; intestine; kidney; pancreas; ss;
 KW metabolic disease; developmental disease; cytostatic; immunomodulatory;
 KW neuroprotective; gene therapy; cancer; immunopathology; neuropathology.
 XX OS Homo sapiens.
 XX PN WO200132927-A2.
 XX PD 10-MAY-2001.
 XX PF 02-NOV-2000; 2000WO-US30396.
 XX PR 04-NOV-1999; 99US-0163508.
 XX PA (INCY-) INCYTE GENOMICS INC.
 XX PI Sornasse T, Seilhamer JJ, Watson GA;
 XX DR WPI: 2001-291057/30.
 XX PT New cell and tissue specific polynucleotides useful for diagnosis,
 PT prognosis or monitoring of treatments for disorders where the gene is
 PT associated with a cancer, immunopathology or neuropathology -
 XX
 PS Claim 1; Page 309-310; 327pp; English.
 XX AAH57161 to AAH57576 represent cell and tissue specific polynucleotide
 CC sequences (I). (I) can have cytostatic, immunomodulatory and
 CC neuroprotective activities and can be used in gene therapy. (I) and
 CC proteins (II) encoded by then are used in high throughput screening
 CC assays to select DNA molecules, RNA molecules, peptide nucleic acids,
 CC mimetics, peptides, proteins, agonists, antagonists, antibodies or
 CC their fragments, immunoglobulins, inhibitors, drug compounds and
 CC pharmaceutical agents. Expression of (I) in a sample indicates the
 CC differentiation of embryonic stem cells into a tissue selected from
 CC brain, heart, kidney, liver, lung, skeletal muscle or pancreatic
 CC tissues. (I) and (II) are used to produce an expression profile that
 CC defines a metabolic or developmental process, treatment, condition,
 CC disease or disorder. The gene profile can be used for diagnosis,
 CC prognosis or monitoring of treatments and for investigating a
 CC predisposition to a disorder where the gene is associated with a
 CC cancer, immunopathology or neuropathology.

CC analysing the expression profile or expression pattern of a gene derived
CC from Corynebacterium, and identifying a homologue of a gene derived
CC from Corynebacterium. Corynebacterium bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
XX Sequence 309400 BP; 70133 A; 86477 C; 83115 G; 69675 T; 0 other;

Query Match 1.9%; Score 49; DB 22; Length 309400;
Best Local Similarity 45.5%; Pred. No. 0.032;
Matches 215; Conservative 0; Mismatches 255; Indels 3; Gaps 1;
QY 473 gatgcatgagatgatgatcctcagcattcctgagccacagatgcatgcatgagtgag 532
DB 227907 gacggtaacacacattgtgtacggcaccgtccaatcgccctgctgaaggagtggaac 227966
QY 533 ctcccaagcggcaggtggcattgctgacccctggtggtctttttaggcattgatgtccagc 592
DB 227967 ttaagctctcaacgctggcgacacatcggtccacgcggttttggcattgcatggcg 228026
QY 593 tccacgctctgggaaatatctcagaccagtcagcgaggagaaacagggctgagatcagc 652
DB 228027 gctgtgtcattgtgtcagctcagacccgctggcgccgaaagcagcggtgattggtacc 228086
QY 653 gtgctgtgagactgtactatgcatctcattgacatttgcgccctggtatagctggtac 712
DB 228087 gtgctgtattctctctctcaccatctgtgtgcatcttctcacaacccatgggtgttc 228146
QY 713 ctggtgctccgggctggtggtggtcctggatcgagagagttcccgatcggtga----cg 769
DB 228147 ggcgttttcggttcatcgacggcctgtgtcggcgctgtgtgcttcagttaatgcg 228206
QY 770 ctgtatcgaggttctctccatgaaagccagagactaaatgtatttgcattgaggtga 829
DB 228207 atgacctctgatttggttccacgttaagaccatgcatcgctgggcaacggtgatgctcg 228266
QY 830 ttcggggccatcgagagtggttcgaggtgctgctgctgctgtgttcgtagtgcacgctg 889
DB 228267 ggcgttcccatcggtgttctatcgcgctgtgtgacacttgggtgttctctctcg 228326
QY 890 ggcgtggcgtggtgctcactcctcagctgtcccgctcctcctcttgcgct 942
DB 228327 gaggagtggtggcgtctcattgttctcctcagctgctgtgacccactgggtgtg 228379

RESULT 10
AAV49559
ID AAV49559 standard; cDNA to mRNA; 1662 BP.
XX
XX AAV49559;
XX
XX 21-OCT-1998 (first entry)
XX
XX Human liver cell clone HP01293 cDNA #2.
XX
XX Transmembrane domain; human; nutrition; cytokine; cell proliferation;
KW differentiation; immune system; stimulator; suppressor; regulator;
KW hematopoiesis; activin; inhibitor; chemotactic; chemokinetic; receptor;
KW haemostatic; thrombolytic; ligand; anti-inflammatory; tumour; ds.
XX
XX Homo sapiens.
XX
XX WO9821328-A2.
XX
XX 22-MAY-1998.
XX
XX 07-NOV-1997; 97WO-JP04056.

PR 13-NOV-1996; 96JP-0301429.
XX
XX (PROT-) PROTEGENE INC.
PA (SAGA) SAGAMI CHEM RES CENTRE.
XX
XX Kato S, Kobayashi M, Sekine S, Yamaguchi T;
PI WPI; 1998-297932/26.
XX
XX P-PSDB; AAW64538.
DR
XX Human protein having transmembrane domain - useful for, e.g.
PT research and nutrition
XX
XX Claim 3; Page 120-121; 205pp; English.
PS
XX AAV49550-V49599 are cDNA sequences which encode human proteins containing
CC a transmembrane domain. These proteins can be used for, e.g. research
CC and nutrition, and may have cytokine and cell
CC proliferation/differentiation, immune stimulating/suppressing,
CC haematopoiesis regulating, tissue growth, activin/inhibin,
CC chemotactic/chemokinetic, haemostatic and thrombolytic, receptor/ligand,
CC anti-inflammatory or tumour inhibition activity.
XX
XX Sequence 1662 BP; 312 A; 486 C; 466 G; 398 T; 0 other;
SQ

Query Match 1.8%; Score 46.6; DB 19; Length 1662;
Best Local Similarity 47.8%; Pred. No. 0.0097;
Matches 232; Conservative 0; Mismatches 244; Indels 9; Gaps 3;
QY 1267 cagcctggcctgcagtlacctgagtgagggaggtattacatgactgtgtggaccacct 1326
DB 1092 catctgcacatggggcgccaccagcggaacctctactgatttctactcgcgtct 1151
QY 1327 ctctgatttctcaggtgctgtgtgacgtctgtgatttattgacgcctggggcgcaagaa 1386
DB 1152 ggtcgaaatccggggcgcttcatagccctcaccattgacgcgtggggcgcatcta 1211
QY 1387 gaccatggcctgtgtgtgtctctctct---cctctgcagcctcctgtgtttattctg 1443
DB 1212 cccatggcgtgctcaaatgttgcgggggcagcctgctcgtcatgattttattctc 1271
QY 1444 tgttggaagaaatgtgctcactctgttactcttca----ttgcaagagcgtttattctgg 1500
DB 1272 acctgacctgcactggttaaacatcataatcatgtgtgtggcgaatgggaatcaccat 1331
QY 1501 agcctttcaagcgcatatgtttacacacctgaggtctaccacacggcaacgcggacct 1560
DB 1332 tgcatacaaatgatctgctgctggtggaatgctgagctgtaccacacattcgtcaggaacct 1391
QY 1561 cggcctgggcacctgcagcgcatggcaagagtggtgtctctcatcactcgtttcatgcg 1620
DB 1392 cggagtgatgtgtgttctctcctctgtgtgacatagggtgggataatcaccccttcagt 1451
QY 1621 ccaggtgatgctggaatcctctgtgtacactgactctggcagtttaccagtgctgctcct 1680
DB 1452 ctccaggtgagggaggtcctgcaagccttgccctcatttctgttgcgggtgtggtg 1511
QY 1681 cctggtgctcctggcctcctctcttcttggccattgagaccacagcgagagctgcagga 1740
DB 1512 gctgctggcggaagtgcgctacttcttcca--gagaccaaggggctgctttgccaaga 1568
QY 1741 gtcca 1745
DB 1569 gacca 1573

RESULT 11
AAV49558
ID AAV49558 standard; cDNA to mRNA; 1888 BP.
XX
XX AAV49558;
XX

```
DT 21-OCT-1998 (first entry)
DE Human liver cell clone HP01293 cDNA #1.
XX
XX Transmembrane domain; human; nutrition; cytokine; cell proliferation;
KW differentiation; immune system; stimulator; suppressor; regulator;
KW haematopoiesis; activin; inhibitor; chemotactic; chemokinetic; receptor;
KW haemostatic; thrombolytic; ligand; anti-inflammatory; tumour; ds.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FT CDS 90..1745
FT /*tag= a
FT /product= "transmembrane domain containing protein"
XX
XX WO9821328-A2.
XX
XX PD 22-MAY-1998.
XX
XX PF 07-NOV-1997; 97WO-JP04056.
XX
XX PR 13-NOV-1996; 96JP-0301429.
XX
XX PA (PROT-) PROTEGENE INC.
XX (SAGA) SAGAMI CHEM RES CENTRE.
XX
XX PI Kato S, Kobayashi M, Sekine S, Yamaguchi T;
DR WPI; 1998-297932/26.
DR P-PSDB; AAW64538.
XX
XX Human protein having transmembrane domain - useful for, e.g.
PT research and nutrition
XX
XX PS Claim 4; Page 141-143; 205pp; English.
XX
XX AAU49550-V49599 are cDNA sequences which encode human proteins containing
CC a transmembrane domain. These proteins can be used for, e.g. research
CC and nutrition, and may have cytokine and cell
CC proliferation/differentiation, immune stimulating/suppressing,
CC haematopoiesis regulating, tissue growth, activin/inhibin,
CC chemotactic/chemokinetic, haemostatic and thrombolytic, receptor/ligand,
CC anti-inflammatory or tumour inhibition activity.
XX
XX SQ Sequence 1888 BP; 365 A; 545 C; 520 G; 458 T; 0 other;

Query Match 1.8%; Score 46.6; DB 19; Length 1888;
Best Local Similarity 47.8%; Pred. No. 0.01;
Matches 232; Conservative 0; Mismatches 244; Indels 9; Gaps 3;

QY 1267 cagcttgccgcgctgacgtgagtgagagattacatggacttgctgtggaccacact 1326
DB ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1181 catctgcacatggcgccaccagcggaacctctacctgacttcccttactccgctct 1240
QY 1327 ctctgagttccaggtgctcttgactctctgtggtattatgacgcctggggcgcaagaa 1386
DB || || || || || || || || || || || || || || || || || || ||
1241 ggtcgaaatccggggcgctctatagccctcatcaccattgacgcgtggcgccatcta 1300
QY 1387 gacctgacctgtgcttctgtctctcttcttcttcttcttcttcttcttcttcttct 1443
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1301 cccatggccgtgtcaaatgttggcgggggcgagctgctgctgcatgattttatctc 1360
QY 1444 ttttgaagaagtgtcactctgttactcttca---ttgcaagagcgctttattcttgg 1500
DB || || || || || || || || || || || || || || || || || || ||
1361 acctgacctgcactggttaaacatcataatcatgtgtgttggcgaatgggaatcacatc 1420
QY 1501 aggtttcaacggcgcataatgtttacacacctgaggtctacccccagcgcaacggggccct 1560
DB || || || || || || || || || || || || || || || || || || ||
1421 tgcatacaaatgatctgcttggcttggatgtcgtgagctgtacccccacattcgtcaggaact 1480
QY 1561 cggctgggacacctgcagcgccggtggaagagtggtgtctctcatcctccttctcatcgc 1620
```

```
DB 1481 cggagtgatggtgtgttctctccctctgtgtgacatagtggggataatcacccctctcatgt 1540
QY 1621 ccaggtgatgctggaatcctctgtgttacctgactctggtgagctttacagtggtgctgcct 1680
DB ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1541 ctccagctgagggaggctggaagccttgcctcatttcttgcggtgttgggctt 1600
QY 1681 cctggtgctccctggtcctctctcttcttcttcttcttcttcttcttcttcttcttctt 1740
DB ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1601 gcttgccgggagtgagcgtactcttcttca---gagaccagggggtgcgttcttgcaga 1657
QY 1741 gtcca 1745
DB 1658 gacca 1662

RESULT 12
AAC93874
ID AAC93874 standard; cDNA; 549 BP.
XX
XX AC AAC93874;
XX
XX DT 19-FEB-2001 (first entry)
XX
XX DE Cat flea hindgut and Malpighian tubule (HMT) cDNA, SEQ ID NO:369.
XX
XX KW Cat flea; hindgut and Malpighian tubule nucleic acid; HMT;
KW flea infestation; vaccine; antiparasitic; therapeutic target;
KW diagnosis; detection; ss.
XX
XX OS Ctenocephalides felis.
XX
XX PN WO200061621-A2.
XX
XX PD 19-OCT-2000.
XX
XX PF 07-APR-2000; 2000WO-US09437.
XX
XX PR 09-APR-1999; 99US-0128704.
XX
XX PA (HESK-) HESKA CORP.
XX
XX PI Brandt KS, Gaines PJ, Stinchcomb DT, Wisniewski N;
XX
XX DR WPI; 2000-656323/63.
XX
XX PT Flea Malpighian tubule and head and nerve cord tissue derived nucleic
PT acids useful for the prevention, diagnosis and treatment of flea
PT infestations
XX
XX PS Claim 26; Page 368; 964pp; English.
XX
XX The invention relates to novel cat flea (Ctenocephalides felis) nucleic
CC acids which are expressed in hindgut and Malpighian tubule (HMT) tissue
CC or head and nerve cord (HNC) tissue. The invention also relates to the
CC encoded proteins. The invention additionally encompasses expression
CC constructs, recombinant viruses and recombinant cells comprising the
CC nucleic acids of the invention, recombinant production of the proteins,
CC antibodies against the proteins, a method of identifying inhibitors of
CC the proteins, and compositions comprising the inhibitors for
CC administration to an animal. The nucleic acids, and the proteins they
CC encode may be used in the prevention, treatment and diagnosis of diseases
CC associated with flea infestations. For example, the nucleic acids may be
CC used to produce an HMT or HNC protein according to standard recombinant
CC DNA methodology by inserting the nucleic acids into a host cell and
CC culturing the cell to express the protein. The HMT and HNC nucleic acids
CC may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect
CC and quantitate the presence of cat flea or other homologous nucleic acid
CC sequences in samples. They may also be used to study the expression and
CC function of the proteins and their role in metabolism. The HMT and HNC
CC proteins may be used as antigens in the production of specific
CC antibodies, and in assays to identify modulators (agonists and
CC antagonists) of HMT and/or HNC protein expression and activity. The
```

or head and nerve cord (HNC) tissue. The invention also relates to the encoded proteins. The invention additionally encompasses expression constructs, recombinant viruses and recombinant cells comprising the nucleic acids of the invention, recombinant production of the proteins, antibodies against the proteins, a method of identifying inhibitors of the proteins and compositions comprising the inhibitors for administration to an animal. The nucleic acids, and the proteins they encode may be used in the prevention, treatment and diagnosis of diseases associated with flea infestations. For example, the nucleic acids may be used to produce an HMT or HNC protein according to standard recombinant DNA methodology by inserting the nucleic acids into a host cell and culturing the cell to express the protein. The HMT and HNC nucleic acids may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect and quantitate the presence of cat flea or other homologous nucleic acid sequences in samples. They may also be used to study the expression and function of the proteins and their role in metabolism. The HMT and HNC proteins may be used as antigens in the production of specific antibodies, and in assays to identify modulators (agonists and antagonists) of HMT and/or HNC protein expression and activity. The anti-HMT/HNC protein antibodies and antagonists may also be used to downregulate protein expression and activity. The antibodies may also be used as diagnostic agents for detecting the presence of flea polypeptides in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The present sequence represents a cat flea HMT cDNA of the invention.

Sequence 1590 BP: 482 A: 252 C: 349 G: 507 T: 0 other:

CC associated with flea infestations. For example, the nucleic acids may be
CC used to produce an HMT or HNC protein according to standard recombinant
CC DNA methodology by inserting the nucleic acids into a host cell and
CC culturing the cell to express the protein. The HMT and HNC nucleic acids
CC may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect
CC and quantitate the presence of cat flea or other homologous nucleic acid
CC sequences in samples. They may also be used to study the expression and
CC function of the proteins and their role in metabolism. The HMT and HNC
CC proteins may be used as antigens in the production of specific
CC antibodies, and in assays to identify modulators (agonists and
CC antagonists) of HMT and/or HNC protein expression and activity. The
CC anti-HMT/HNC protein antibodies and antagonists may also be used to
CC downregulate protein expression and activity. The antibodies may also be
CC used as diagnostic agents for detecting the presence of flea polypeptides
CC in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The
CC present sequence represents a cat flea HMT cDNA of the invention.
XX
SQ Sequence 1590 BP; 482 A; 252 C; 349 G; 507 T; 0 other;

Query Match 1.8%; Score 46.4; DB 21; Length 1590;
Best Local Similarity 48.8%; Pred. No. 0.011;
Matches 125; Conservative 0; Mismatches 131; Indels 0; Gaps 0

Qy	391	ggtggaaagatgcagtgggaagccattgctttggaaaatttcctcagtgaggacgtctctgttct	450
Db	57	gttgaagatcgcttggcactcacaggttttggaaaattcaattacctctctctggtcgt	116
Qy	451	cactggcttggcttggatggcttgatgccatggagatgatcctcagcatcctgcacc	510
Db	117	gagtgagtcatttagtatgttttcatggaacctcttggaatgagtttgtcgatccc	176
Qy	511	acagctgcatgcagtgaggagctccccagctcggcaggtggcattgctgcacctcgglggt	570
Db	177	ttcagcaaatgttgatctggaaataaaccaaaaaacaaaaggaaatttaagcgtctatagc	236
Qy	571	ctttgtaggcatgatgtccagctccacgctctcggggaaatatctcagaccagctacggcag	630
b	237	ttttatagcatatttaaacagcttgcacattttatggagaattttttacagcatatcacaagagag	296

Qy	631	gaaacacaggctgaag	646
Db	297	gcgaaagtgtattg	312
RESULT	14		
AAC95402/C			
ID	AAC95402	standard; cDNA;	1590 BP.
XX	XX		
AC	AAC95402;		
XX	XX		
DT	19-FEB-2001	(first entry)	
XX	XX		
DE	Cat flea HMT synaptic vesicle 2B-like cDNA complement,	SEQ ID NO:1905.	
XX	XX		
KW	Cat flea; hindgut and Malpighian tubule nucleic acid;	HMT;	
KW	flea infestation; vaccine; antiparasitic; therapeutic target;		
KW	diagnosis; detection; ss.		
XX	XX		
OS	Ctenocephalides felis.		

KW	flea infestation; vaccine; antiparasitic; therapeutic target;
KW	diagnosis; detection; ss.
XX	
XX	Ctenocephalides felis.
OS	
XX	WO200061621-A2.
XX	
PD	19-OCT-2000.
XX	

```
PF 07-APR-2000; 2000WO-US09437.
PR 09-APR-1999; 99US-0128704.
XX (HESK-) HESKA CORP.
XX Brandt KS, Gaines PJ, Stinchcomb DT, Wisniewski N;
XX WPI; 2000-656323/63.
DR P-PSDB; AAB29626.
XX Flea Malpighian tubule and head and nerve cord tissue derived nucleic
PT acids useful for the prevention, diagnosis and treatment of flea
PT infestations.
XX Claim 1; Page 927; 964pp; English.
XX The invention relates to novel cat flea (Ctenocephalides felis) nucleic
CC acids which are expressed in hindgut and Malpighian tubule (HMT) tissue
CC or head and nerve cord (HNC) tissue. The invention also relates to the
CC encoded proteins. The invention additionally encompasses expression
CC constructs, recombinant viruses and recombinant cells comprising the
CC nucleic acids of the invention, recombinant production of the proteins,
CC antibodies against the proteins, a method of identifying inhibitors of
CC the proteins, and compositions comprising the inhibitors for
CC administration to an animal. The nucleic acids, and the proteins they
CC encode may be used in the prevention, treatment and diagnosis of diseases
CC associated with flea infestations. For example, the nucleic acids may be
CC used to produce an HMT or HNC protein according to standard recombinant
CC DNA methodology by inserting the nucleic acids into a host cell and
CC culturing the cell to express the protein. The HMT and HNC nucleic acids
CC may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect
CC and quantitate the presence of cat flea or other homologous nucleic acid
CC sequences in samples. They may also be used to study the expression and
CC function of the proteins and their role in metabolism. The HMT and HNC
CC proteins may be used as antigens in the production of specific
CC antibodies, and in assays to identify modulators (agonists and
CC antagonists) of HMT and/or HNC protein expression and activity. The
CC anti-HMT/HNC protein antibodies and antagonists may also be used to
CC downregulate protein expression and activity. The antibodies may also be
CC used as diagnostic agents for detecting the presence of flea polypeptides
CC in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The
CC present sequence represents a cat flea HMT cDNA of the invention.
XX Sequence 1590 BP; 507 A; 349 C; 252 G; 482 T; 0 other;
SQ
Query Match 1.8%; Score 46.4; DB 21; Length 1590;
Best Local Similarity 48.8%; Pred. No. 0.011;
Matches 125; Conservative 0; Mismatches 131; Indels 0; Gaps 0;
QY 391 ggtggaagatgcagtggaagccattggttggaaaatttcagtggaagctgtcttct 450
DB 1534 GTTTGAAGATGCGTGGGCACCTACAGGTTTGGAAATTCAAATACCTTCTTCGTGCTGT 1475
QY 451 cactggctggcttggtggtgatgccatggagatgatcctcagcatcctgcacc 510
DB 1474 GAGTGGATGCGTATTAGTATGTTTGTGATGAAACTCTTGGAAATGAGTTTGTGCTTC 1415
QY 511 acagctgcattgcagtgagggtcccaagctgcagctggtgcattgctgactcgtggt 570
DB 1414 TTCAGCACAAATGTGATCTGGAATTAACACAAACAAAGGAATATTTAAGCGCTATAGC 1355
QY 571 cttttaggcattgattccagctccacgctctgtgggaaatactcagaccagtcagcag 630
DB 1354 TTTTATAGTATTATTAAGCAGTTTACATTTATTTGCGGATTTTACCGGATACGAGGGAG 1295
QY 631 gaaacacagggtgaag 646
DB 1294 GCGGAAAGTGATTATG 1279
RESULT 15
AAC95399
ID AAC95399 standard; cDNA; 1875 BP.
XX
AC AAC95399;
XX
DT 19-FEB-2001 (first entry)
XX
DE Cat flea HMT synaptic vesicle 2B-like cDNA, SEQ ID NO:1901.
XX
KW Cat flea; hindgut and Malpighian tubule nucleic acid; HMT;
KW flea infestation; vaccine; antiparasitic; therapeutic target;
KW diagnosis; detection; ss.
XX
OS Ctenocephalides felis.
XX
PN WO200061621-A2.
XX
PD 19-OCT-2000.
XX
PF 07-APR-2000; 2000WO-US09437.
XX
PR 09-APR-1999; 99US-0128704.
XX (HESK-) HESKA CORP.
XX
PI Brandt KS, Gaines PJ, Stinchcomb DT, Wisniewski N;
XX
DR WPI; 2000-656323/63.
DR P-PSDB; AAB29626.
XX
PT Flea Malpighian tubule and head and nerve cord tissue derived nucleic
PT acids useful for the prevention, diagnosis and treatment of flea
PT infestations.
XX
PS Claim 1; Page 920-923; 964pp; English.
XX The invention relates to novel cat flea (Ctenocephalides felis) nucleic
CC acids which are expressed in hindgut and Malpighian tubule (HMT) tissue
CC or head and nerve cord (HNC) tissue. The invention also relates to the
CC encoded proteins. The invention additionally encompasses expression
CC constructs, recombinant viruses and recombinant cells comprising the
CC nucleic acids of the invention, recombinant production of the proteins,
CC antibodies against the proteins, a method of identifying inhibitors of
CC the proteins, and compositions comprising the inhibitors for
CC administration to an animal. The nucleic acids, and the proteins they
CC encode may be used in the prevention, treatment and diagnosis of diseases
CC associated with flea infestations. For example, the nucleic acids may be
CC used to produce an HMT or HNC protein according to standard recombinant
CC DNA methodology by inserting the nucleic acids into a host cell and
CC culturing the cell to express the protein. The HMT and HNC nucleic acids
CC may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect
CC and quantitate the presence of cat flea or other homologous nucleic acid
CC sequences in samples. They may also be used to study the expression and
CC function of the proteins and their role in metabolism. The HMT and HNC
CC proteins may be used as antigens in the production of specific
CC antibodies, and in assays to identify modulators (agonists and
CC antagonists) of HMT and/or HNC protein expression and activity. The
CC anti-HMT/HNC protein antibodies and antagonists may also be used to
CC downregulate protein expression and activity. The antibodies may also be
CC used as diagnostic agents for detecting the presence of flea polypeptides
CC in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The
CC present sequence represents a cat flea HMT cDNA of the invention.
XX Sequence 1590 BP; 507 A; 349 C; 252 G; 482 T; 0 other;
SQ
Query Match 1.8%; Score 46.4; DB 21; Length 1875;
Best Local Similarity 48.8%; Pred. No. 0.012;
Matches 125; Conservative 0; Mismatches 131; Indels 0; Gaps 0;
QY 391 ggtggaagatgcagtggaagccattggttggaaaatttcagtggaagctgtcttct 450
DB 100 gtttgaagatgcgttggcactcacaggttttggaaaatttcattaccttcttctgtcgt 159
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 7, 2002, 23:45:17 ; Search time 137.14 Seconds
(without alignments)
4230.980 Million cell updates/sec

Title: US-09-911-667A-1
Perfect score: 2562
Sequence: 1 cccacgcgtccggcttctgt.....tgaggactgagngccaagtq 2562

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues
Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCUTS.COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53.2	2.1	1885	3	US-08-501-572-5
2	53.2	2.1	1885	3	US-09-040-444-5
3	45.2	1.8	7218	1	US-08-232-463-14
4	45	1.8	1896	3	US-08-501-572-6
5	45	1.8	1896	3	US-09-040-444-6
6	42	1.6	7218	1	US-08-232-463-14
7	38.6	1.5	1882	3	US-08-501-572-4
8	38.6	1.5	1882	3	US-09-040-444-4
9	37.6	1.5	50937	4	US-09-428-517-1
10	36.6	1.4	1931	2	US-09-130-114-2
11	36.4	1.4	1490	4	US-08-964-127-3
12	36.4	1.4	2460	4	US-08-964-127-1
13	35.6	1.4	3489	2	US-08-728-323A-1
14	35.6	1.4	32207	2	US-08-770-379-20
15	35.6	1.4	32207	4	US-08-757-669A-20
16	35.2	1.4	2496	1	US-08-073-384C-2
17	35.2	1.4	2496	1	US-08-254-359A-2
18	35.2	1.4	2496	1	US-08-483-043-2
19	35.2	1.4	2496	1	US-08-481-238-2
20	35.2	1.4	2496	2	US-08-471-066B-2
21	35.2	1.4	2496	2	US-08-484-956-2
22	35.2	1.4	2496	2	US-08-757-653-2
23	35.2	1.4	2496	2	US-08-599-491-2
24	35.2	1.4	2496	2	US-08-756-386-2
25	35.2	1.4	2496	2	US-08-823-516-2
26	35.2	1.4	2496	3	US-08-682-853A-2
27	35.2	1.4	2496	3	US-08-759-038-2

28	35.2	1.4	2496	3	US-08-758-314-2	Sequence 2, Appli
29	35	1.4	2082	4	US-09-440-325A-2	Sequence 2, Appli
30	34.8	1.4	2102	2	US-08-647-397-1	Sequence 1, Appli
31	34.4	1.3	289	4	US-09-007-005-17	Sequence 17, Appli
32	34.4	1.3	289	4	US-09-244-796-17	Sequence 17, Appli
33	34.2	1.3	2484	4	US-09-276-531-46	Sequence 46, Appli
34	34	1.3	2515	1	US-08-061-465-3	Sequence 3, Appli
35	34	1.3	6828	1	US-08-061-465-1	Sequence 1, Appli
36	33.6	1.3	803	1	US-07-928-611-12	Sequence 12, Appli
37	33.6	1.3	803	2	US-08-487-811A-12	Sequence 12, Appli
38	33.6	1.3	803	4	US-09-060-694-12	Sequence 12, Appli
39	33.6	1.3	803	5	PCT-US93-07370-12	Sequence 5, Appli
40	33.6	1.3	1610	1	US-08-056-051-5	Sequence 21, Appli
41	33.6	1.3	1610	1	US-07-928-611-21	Sequence 21, Appli
42	33.6	1.3	1610	2	US-08-487-811A-21	Sequence 21, Appli
43	33.6	1.3	1610	4	US-09-060-694-21	Sequence 21, Appli
44	33.6	1.3	1610	5	PCT-US93-07370-21	Sequence 21, Appli
45	33.6	1.3	2019	4	US-09-063-950-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-501-572-5
; Sequence 5, Application US/08501572
; Patent No. 6063623
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; TITLE OF INVENTION: Transport protein Which Effects The
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics and/or Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan,Henderson,Farabow,Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/501.572
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Toohey, Kimberlin M
; REGISTRATION NUMBER: 35,391
; REFERENCE/DOCKET NUMBER: 02481.1453-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1885 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-501-572-5

Query Match 2.1%; Score 53.2; DB 3; Length 1885;
Best Local Similarity 47.3%; Pred. No. 1e-05;
Matches 228; Conservative 0; Mismatches 248; Indels 6; Gaps 2;
Oy 1267 cagcctgcgcgagctacctgagtgaggagattacatggacttctgtggaccacct 1326

```
Db 1164 CATCTGACATGGCGGCACACAGGGAACCTCTACCTGGATTTCCTTTACTCCGCTCT 1223
Qy 1327 clctgagttccagtgctctgtgactctgtgattattgaccgctgggagcgcaagaa 1386
Db 1224 GGTGAAATCCCGGGGCTTCATAGCCCTCATACCATTTGACCGCGTGGCGGCATCTA 1283
Qy 1387 gaccatggccctgtgcttctgtatctctctctctctctctctctctctctctctct 1446
Db 1284 CCCCATGGCGGTGCAAAATTTGTTGGCGGGGACAGCCTGCCCTCGTCATTTTATCTCACC 1343
Qy 1447 tgaagaagaatgctcactctgtactcttca---ttcaagaagcgtttatttctggagg 1503
Db 1344 TGACCTGCACTGGTTAAACATCATATCATGTGTGTGGCCGAATGGAATCACCATTGC 1403
Qy 1504 cttcaagcggcagatgttttacacacctgaggtctacccccagcgcaacgcgccctcgg 1563
Db 1404 AATACAAATGATCTGCCTGGTGAATGCTGAGCTGTACCCACATTCGTCAGGAACCTCAG 1463
Qy 1564 cctgggacacctcagcgagcagtggaagagtggtgctctctctctctctctctctctc 1623
Db 1464 AGTGATGGGTGTTCTCCCTCTGTGACATAGGTGGGATAATCACCCCTTCATATAGTCTT 1523
Qy 1624 ggtgatgctggaatcctctgtgactgactctgagctgtgacagtggtcgtcctcct 1683
Db 1524 CAGGCTGAGGAGGTCTGCGAAGCCTTGCCCTCATTTTGTGGGTGTTGGGCTGCT 1583
Qy 1684 ggtctgctgctcctctctctctctctctctctctctctctctctctctctctctc 1743
Db 1584 TGCCCGGGAGGTGAGCGCTACTTCTTCC---AGAGACCAAGGGGGACGCTTTGCCAGAGAC 1640
Qy 1744 ca 1745
Db 1641 CA 1642
```

RESULT 2

```
US-09-040-444-5
; Sequence 5, Application US/09040444
; Patent No. 6063766
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grudeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein which effects the
; TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,444
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
```

```
; LENGTH: 1885 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-040-444-5
```

```
Query Match 2.1%; Score 53.2; DB 3; Length 1885;
Best Local Similarity 47.3%; Pred. No. 1e-05;
Matches 228; Conservative 0; Mismatches 248; Indels 6; Gaps 2;
```

```
Qy 1267 cagcctggcctgcgagctacgtgagtgagagattacatggaactgctgctgagcaaccct 1326
Db 1164 CAICCTGCACATGGCGGCACACAGGGAACCTCTACCTGGATTTCCTTTACTCCGCTCT 1223
Qy 1327 clctgagttccagtgctctgtgactctgaggtatttaccgctgggagcgcaagaa 1386
Db 1224 GGTGAAATCCCGGGGCTTCATAGCCCTCATACCATTTGACCGCGTGGCGGCATCTA 1283
Qy 1387 gaccatggccctgtgcttctgtatctctctctctctctctctctctctctctctct 1446
Db 1284 CCCCATGGCGGTGCAAAATTTGTTGGCGGGGACAGCCTGCCCTCGTCATTTTATCTCACC 1343
Qy 1447 tgaagaagaatgctcactctgtactcttca---ttcaagaagcgtttatttctggagg 1503
Db 1344 TGACCTGCACTGGTTAAACATCATATCATGTGTGTGGCCGAATGGAATCACCATTGC 1403
Qy 1504 cttcaagcggcagatgttttacacacctgaggtctacccccagcgcaacgcgccctcgg 1563
Db 1404 AATACAAATGATCTGCCTGGTGAATGCTGAGCTGTACCCACATTCGTCAGGAACCTCAG 1463
Qy 1564 cctgggacacctcagcgagcagtggaagagtggtgctctctctctctctctctctctc 1623
Db 1464 AGTGATGGGTGTTCTCCCTCTGTGACATAGGTGGGATAATCACCCCTTCATATAGTCTT 1523
Qy 1624 ggtgatgctggaatcctctgtgactgactctgagctgtgacagtggtcgtcctcct 1683
Db 1524 CAGGCTGAGGAGGTCTGCGAAGCCTTGCCCTCATTTTGTGGGTGTTGGGCTGCT 1583
Qy 1684 ggtctgctgctcctctctctctctctctctctctctctctctctctctctctctc 1743
Db 1584 TGCCCGGGAGGTGAGCGCTACTTCTTCC---AGAGACCAAGGGGGACGCTTTGCCAGAGAC 1640
Qy 1744 ca 1745
Db 1641 CA 1642
```

RESULT 3

```
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
```

FILING DATE: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/501,572
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Toohy, Kimberlin M
REGISTRATION NUMBER: 35,391
REFERENCE/DOCKET NUMBER: 02481.1453-00000
TELEPHONE: (202)408-4000
TELEFAX: (202)408-4400
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1896 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
CLONE: ptz9pt-Fls
US-08-232-463-14

Query Match 1.88; Score 45.2; DB 1; Length 7218;
Best Local Similarity 4.08; Pred. No. 0.0053;
Matches 14; Conservative 194; Mismatches 142; Indels 0; Gaps 0;

Qy 1365 ttgaccctggcgcaagaagaccatgcccctgtgttgcattctctctctgca 1424
Db 1039 TTGCTGAGGTCGAGGAGCTTCCGATYYYYYYYYYYYYYYYYYYYY 1098
Qy 1425 gctcctgctgttattgtgttgaagaatgtgctcactgttactcttca 1484
Db 1099 YY 1158
Qy 1485 gacgttattctggaggtttcaagcgcatatgtttacacactgaggtc 1544
Db 1159 YY 1218
Qy 1545 cggcaacgcgccctgctgctgagcactgagcgcatggaagagtgctc 1604
Db 1219 YY 1278
Qy 1605 tcactcgttcatcgccaggtgatgctggaatcctctgttacctgactg 1664
Db 1279 YY 1338
Qy 1665 acagtggcgtcctgctgctgctgctgctgcttcttggccatt 1714
Db 1339 YY 1388

RESULT 4
US-08-501-572-6
Sequence 6, Application US/08501572
Patent No. 6063623
GENERAL INFORMATION:
APPLICANT: Koepsell, Hermann
APPLICANT: Grundeman, Dirk
TITLE OF INVENTION: Transport protein which effects the
TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
TITLE OF INVENTION: DNA sequences encoding it and their use.
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA

Qy 1290 gtaaggaggattacgactgctgtggaccacccctctgagttccagtgctctg 1349
Db 1362 GTGCAATATCTACTGGATTTCTTCTACTCTGCCGTGGTTGAATCCGACTGCCTCA 1321
Qy 1350 tgactctgtgattatgaccgctggggcgcaagaagaccatggc-----cctgtgctt 1404
Db 1322 TGATCATCTCATTTATCGACCGCATCGGACGCGTTACCTTGGGCTGCATCAATATGG 1381
Qy 1405 tgcattctctctgagcctcctgcttcttctctctctctctctctctctctctctct 1463
Db 1382 TTGCAGGGGAGCGCTGTCTGGCCCTCAGTTTTTATACCTGGTGTGATCTACAAATGGCTAAAA 1441
Qy 1464 ctctgttactctctctgcaagagcgtttattcttctggagccttctcaagcgcatatgttt 1523
Db 1442 TTATTATCTCATGCTGGGAAGATGGGATCAACAATGGCCCTATGAGATAGTCTGCCTGG 1501
Qy 1524 acacactgaggtctaccccaacgcaacgcccctggcctggcctggcctggcctggcctggc 1583
Db 1502 TCAATGCTGAGCTGACCCCATTCATTAGGAATCTTGGCGTCCACATCTGTTCCTCAA 1561
Qy 1584 tggcaagagtggtgctctcactcctcctcctcctcctcctcctcctcctcctcctc 1643
Db 1562 TGTGTGACATGGTGGCATCATCAGCCATTCCTGGTGTACGGCTCACTACATCTGCG 1621
Qy 1644 tgcactgactctggcaggtttacagtggtgctgctcctcctcctcctcctcctcctc 1688
Db 1622 TTGAGCTCCCGCTGATGTTTTTCGGCGTACTTGGCTTGGTGTGCTG 1666

RESULT 5
US-09-040-444-6
Sequence 6, Application US/09040444
Patent No. 6063766
GENERAL INFORMATION:
APPLICANT: Koepsell, Hermann
APPLICANT: Grundeman, Dirk
APPLICANT: Gorboulev, Valentin
TITLE OF INVENTION: Transport protein which effects the
TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
TITLE OF INVENTION: DNA sequences encoding it and their use.
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:

```
; ADDRESSEE: Finnegan,Henderson,Parabow,Garrett & Dunner, L.L.P.
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040.444
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1896 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-040-444-6

Query Match 1.8%; Score 45; DB 3; Length 1896;
Best Local Similarity 47.9%; Pred. No. 0.0028;
Matches 194; Conservative 0; Mismatches 205; Indels 6; Gaps 2;

QY 1290 gtgaggagattacatggacttgctgtagcaccctctctcagtttcacagtgctctg 1349
DB 1262 GTGACAAATATCTACCTGGATTTCTTCTACTCTGCGCTGTTGAATTCACAGCTGCTTCA 1321

QY 1350 tgactctgtgattatgaccgcctggggcgcaagaagaccatggc-----cctgtgctt 1404
DB 1322 TGATCATCTCATGCTTGGGAAGAAATGGGATCACAAATGGCCATGATGATGCTCAAAATATGG 1381

QY 1405 tgcattctctctcagcagctcctgctgtttatct-gtcttggaagaagaatgctca 1463
DB 1382 TTGACGGGACGCTCTGCGCTCTAGTTTTTATACCTGGTGATCTACAATGGCTAAAAA 1441

QY 1464 ctctgttactcttcattgcaagagcgtttatttctggaggccttcaagcggcatatgttt 1523
DB 1442 TTATTATCTCATGCTTGGGAAGAAATGGGATCACAAATGGCCATGATGATGCTGCTGG 1501

QY 1524 acacactgagcttaccacagcgaacgcggccctcgccctgggcaacctgcagcgga 1583
DB 1502 TCAACTCTGAGCTGTACCCCAATTCATTAGGAATCTTGGCTGCCACATCTCTCTCTCAA 1561

QY 1584 tggcaagagtgggtgctctcactcccttcacgcagagtgatgctgggaatcctctg 1643
DB 1562 TGTGTGACATTTGGTGACATCATCACGCCATTCCTGTGCTACCGGCTCACTAATCATCTGCG 1621

QY 1644 tglactgactctggcagtttcagtgctgctgctcctcctgctg 1688
DB 1622 TTGAGCTCCCGTATGGGTTTTTGGCGGTACTTGGCTTGTGCTG 1666
```

RESULT 6

```
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
```

```
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match 1.6%; Score 42; DB 1; Length 7218;
Best Local Similarity 4.1%; Pred. No. 0.048;
Matches 15; Conservative 200; Mismatches 155; Indels 0; Gaps 0;

QY 146 aaggcgcatgagaagggctcactcctggcgacgcgcagatgagagagacttattccag 205
DB 1411 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1352

QY 206 ctaaggcagctcccggttgtgaaattccctgcacagcgagagtgcaaggtcagaggac 265
DB 1351 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1292

QY 266 gacacggcttcaggagagcatgaagtcacagattgaaggggtccacgtggccttagagct 325
DB 1291 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1232

QY 326 gggagctgagatggggcagctgtgccaaaggatttgccaatcccaccatgatact 385
DB 1231 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1172

QY 386 ttcatggtggaagatgcagtgaaagccattgctttggaaaaatttcagtggaagctgtct 445
DB 1171 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1112

QY 446 gtctcactggttggttggtgctgctgacatgagatgatgatcctcagcatcctg 505
DB 1111 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1052

QY 506 gcaccacagc 515
DB 1051 GACCTGCACC 1042
```

RESULT 7

US-08-501-572-4
; Sequence 4, Application US/08501572
; Patent No. 6063623
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein Which Effects The
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics and/or Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan,Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/501.572
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Toohy, Kimberlin M
; REGISTRATION NUMBER: 35,391
; REFERENCE/DOCKET NUMBER: 02481.1453-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1882 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-501-572-4

Query Match 1.5%; Score 38.6; DB 3; Length 1882;
Best Local Similarity 47.0%; Pred. No. 0.22;
Matches 224; Conservative 0; Mismatches 244; Indels 9; Gaps 3;

OY 1301 tacatggacttgcgtggaccaccctctcagttccaggtgctctgtgactctgtgg 1360
DB 1166 TACCTGGACTCTTTTATTCTCTGTTGGAATTCCTCCCGGGCTTCATCATCTGTGC 1225
OY 1361 attatgaccgcctggggcgcaagaagaccat---ggccctgtgcttgcattctctcc 1417
DB 1226 ACCATTGACCGCATTTGGCGCATCTACCCAATAGCGGCTCGAATCTGTGACGGGGCA 1285
OY 1418 ttctcagccctcctgttctatctgttggagaagaatgtcctcactgttact---c 1474
DB 1286 GCCTGCCTCCTCATGATCTTTATCCCGCATGAGCTGCACCTGGTTGAACCTTACCTCGCC 1345
OY 1475 ttcatgcaagcgcttattcttctgagggtcttcaagcgccatattttacacactgag 1534
DB 1346 TGCTCTGGCCGTATGGGGCCACCATTGTGCTGCAGATGGTCTGCTGGTGAACCTGAG 1405
OY 1535 gtctacccacggcaacgcggccctgcctggccactcagcgcatggcaagaagt 1594
DB 1406 CTGTACCTTACATTCATCAGGAATCTTGGGATGATGGATGCTCTGCCCTGTGTGACCTG 1465
OY 1595 ggtctctcactcctccgtttcctatgcgccagggtgatgctgggaatcctctgtgtacctgact 1654
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 1466 GGTGGATCTTCACCCCTTCATGCTGTTCAGGCTGATGGAAGTTTGGCAAGCCCTGCOC 1525
OY 1655 ctggcagtttacagtggctgctgctcctgctgctgcctgcctgcctcctgttttgcacatt 1714
DB 1526 CT---CATTTTGTGGGTTTGGGCTGACTGTGGGGCATGACTCTTCTTCTCCCA 1582
OY 1715 gagacaaaagcgagagactcagagagtcagagccagccagggagtgggccagagagt 1771
DB 1583 GAGACCAAGGGTGTGCTTTCCTGAGACTATTGAAGAAGCAGAGAACCTGGGGAGG 1639
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 8

US-09-040-444-4
; Sequence 4, Application US/09040444
; Patent No. 6063766
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein Which Effects The
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics and/or Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan,Henderson, Farabow, Garrett & Dunner, L.L.P.
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/040.444
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1882 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-040-444-4

Query Match 1.5%; Score 38.6; DB 3; Length 1882;
Best Local Similarity 47.0%; Pred. No. 0.22;
Matches 224; Conservative 0; Mismatches 244; Indels 9; Gaps 3;

OY 1301 tacatggacttgcgtggaccaccctctcagttccaggtgctctgtgactctgtgg 1360
DB 1166 TACCTGGACTCTTTTATTCTCTGTTGGAATTCCTCCCGGGCTTCATCATCTGTGC 1225
OY 1361 attatgaccgcctggggcgcaagaagaccat---ggccctgtgcttgcattctctcc 1417
DB 1226 ACCATTGACCGCATTTGGCGCATCTACCCAATAGCGGCTCGAATCTGTGACGGGGCA 1285
OY 1418 ttctcagccctcctgttctatctgttggagaagaatgtcctcactgttact---c 1474
DB 1286 GCCTGCCTCCTCATGATCTTTATCCCGCATGAGCTGCACCTGGTTGAACCTTACCTCGCC 1345
OY 1475 ttcatgcaagcgcttattcttctgagggtcttcaagcgccatattttacacactgag 1534
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1346 TGCTTGCCGATATGGGGCCACCATTTGCTGCAGATGGTCTGCCGTGGTGAACCTGAG 1405
Qy 1535 gtctacccacgcaacgcgccctcgccctggcgacactgcagcgccatggcaagagt 1594
Db 1406 CTGTACCTACATTCATCAGAAATCTGGGATGATGGATTGCTCTGCCCTGTGTGACCTG 1465
Qy 1595 ggtctctcatcatcctcggttcacgcgccaggtgtagctggaatcctctgtgtacctgact 1654
Db 1466 GGTGGGATCTTACCCCTTTCAGGCTGTTCAGGCTGATGGAAGTTTGGCAAGCCCTGCC 1525
Qy 1655 ctggcagttacagtggtctgctctgctcctgctcctcctcctcctcctcctcctcctcatt 1714
Db 1526 CT---CATTTGTTGGGTTTGGGCTTGGGCTGACTGCTGGGGCCATGACTCTTCTTCTCCCA 1582
Qy 1715 gacacaaaggcgaggactcagagtgccagccaccgggagtgggggcagagatg 1771
Db 1583 GAGACCAAGGTTGGCTTTGCTCTGAGACTATTGAAGAAGCAGACAACTGGGGAGG 1639
RESULT 9
US-09-428-517-1
; Sequence 1, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDuniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/428,517
; CURRENT FILING DATE: 1999-10-28
; EARLIER FILING DATE: 1999-02-16
; EARLIER APPLICATION NUMBER: 60/106,100
; EARLIER FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA
US-09-428-517-1
Query Match 1.5%; Score 37.6; DB 4; Length 50937;
Best Local Similarity 54.3%; Pred. No. 3;
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
Qy 255 ggtcagagcagacacgcttcaggagagcagatgaagtcagattgaagggtccacgtgg 314
Db 37483 ggttgacggttcgcgcgcacatcgtctgcgaatggcgttgacgtgagcgagccatga 37542
Qy 315 gctcagagcgttgagctggatgatgaggcgagctgtgcccagagagtttgcaatccca 374
Db 37543 tctgggacgcggtcgagcgtgaggaactcgtcgtcgttagccgaggagttggggat-gacg 37602
Qy 375 ccgatatactttcattgg 394
Db 37603 ccgagcgttgacacacggtg 37622
RESULT 10
US-09-130-114-2/c
; Sequence 2, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damaj, Bassam B.
; APPLICANT: Robbins, Alan K.

; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; TITLE OF INVENTION: From Multiple Transfected Episomes
; FILE REFERENCE: 0867/1D903US1
; CURRENT APPLICATION NUMBER: US/09/130,114
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1931
; TYPE: DNA
; ORGANISM: EBNA
; US-09-130-114-2
Query Match 1.4%; Score 36.6; DB 2; Length 1931;
Best Local Similarity 46.1%; Pred. No. 0.88;
Matches 123; Conservative 0; Mismatches 144; Indels 0; Gaps 0;
Qy 146 aaggcgcatgaaaggctcctcctgcgagcgcgacatggaggaggacttattccag 205
Db 1078 AAGTGCAGGAGATGGGGCCCGCGAGGTGATGGAGGAGCTGGGGCCCGCGAGGTGATGGAG 1019
Qy 206 ctaaggcagctgcgggttgtaaaatccgtcgcacagcgagagtgcaagggtcagagac 265
Db 1018 GAGCTGGGGCCCGAGGTGACGAGGAGCTGGGGCCCGAGGTGAGGAGGACGACGGGAG 959
Qy 266 gacacggttcaggagagcagatgaagtcagattgaagggtccacgtggcctagaggt 325
Db 958 GACGAGGACGGGAGGAGGACGAGGACGGGAGGACGGGAGGACGAGGACGGGAGGAC 899
Qy 326 gtgagctggatgagggcgagctgtgcccaggagttgccaatcccaatcccaatgatact 385
Db 898 GGGGAGGACGAGGACGGGAGGACGGGAGGACGAGGACGGGAGGACGGGAGGAGGAC 839
Qy 386 tL..tggTggaagatgcagtgaagcc 412
Db 838 GAGGACGGGAGGACGGGAGGAGGAC 812
RESULT 11
US-08-964-127-3
; Sequence 3, Application US/08964127
; Patent No. 627565
; GENERAL INFORMATION:
; APPLICANT: Grandearl, Andrew David John
; TITLE OF INVENTION: NOVEL GENES ENCODING TRANSPORTER-LIKE
; TITLE OF INVENTION: MOLECULES
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/964,127
; FILING DATE: 06-NOV-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Crews, Ph.D., L. Lee
; REGISTRATION NUMBER: P-43,567
; REFERENCE/DOCKET NUMBER: 07334/038001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906

TELEX: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1490 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 492...1349
US-08-964-127-3

Query Match 1.4%; Score 36.4; DB 4; Length 1490;
Best Local Similarity 47.1%; Pred. No. 0.86;
Matches 112; Conservative 0; Mismatches 126; Indels 0; Gaps 0;
Qy 535 cccaagtcgagtgagcattgctgacccggtggtctttagcagcatgattccagctc 594
Db 776 CTGGGCTGGCAGGTGATCTTGGAGCAGATCCTCTTATCTTGGGCTTTGGCTCGGCTA 835
Qy 595 cagctctggggaataatctcagaccagtcagggcaggaacagggctgaagatcagcgt 654
Db 836 CTTGTTCTGGTGTACCCCGCAGACAGATTTGGCCGTCGCGGATTTGCTGCTGACCTT 895
Qy 655 gctgtgagctctgactatgagcattccttagtgatttgcgcccgtgtatagctgatact 714
Db 896 GGGGCTGTGGGCCCTGTGGAGTAGGAGGGGCTGCTGCAGGGCTCCCTCCACAGCGTCTAT 955
Qy 715 ggtgtccggggcctggtggcctcgggagtcggaggagttcccccagtcggtgacgctg 772
Db 956 GGCCTCCGATTCCTTTGGGCTTTCTGCTTGGCGGTGTTGACCTGGGTGCTACCTG 1013

RESULT 12
US-08-964-127-1
Sequence 1, Application US/08964127
Patent No. 6277565
GENERAL INFORMATION:
APPLICANT: Grandearl, Andrew David John
TITLE OF INVENTION: NOVEL GENES ENCODING TRANSPORTER-LIKE
MOLECULES
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/964.127
FILING DATE: 06-NOV-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Crews, Ph.D., L. Lee
REGISTRATION NUMBER: P-43,567
REFERENCE/DOCKET NUMBER: 07334/038001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2460 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 498...2057
US-08-964-127-1

Query Match 1.4%; Score 36.4; DB 4; Length 2460;
Best Local Similarity 47.1%; Pred. No. 1.2;
Matches 112; Conservative 0; Mismatches 126; Indels 0; Gaps 0;
Qy 535 cccaagtcgagtgagcattgctgacccggtggtctttagcagcatgattccagctc 594
Db 782 CTGGGCTGGCAGGTGATCTTGGAGCAGATCCTCTTATCTTGGGCTTTGGCTCGGCTA 841
Qy 595 cagctctggggaataatctcagaccagtcagggcaggaacagggctgaagatcagcgt 654
Db 842 CTTGTTCTGGTGTACCCCGCAGACAGATTTGGCCGTCGCGGATTTGCTGCTGACCTT 901
Qy 655 gctgtgagctctgactatgagcattccttagtgatttgcgcccgtgtatagctgatact 714
Db 902 GGGGCTGTGGGCCCTGTGGAGTAGGAGGGGCTGCTGCAGGCTCCTCCACAGCGTCTAT 961
Qy 715 ggtgtccggggcctggtggcctcgggagtcggaggagttcccccagtcggtgacgctg 772
Db 962 GGCCTCCGATTCCTTTGGGCTTTCTGCTTGGCGGTGTTGACCTGGGTGCTACCTG 1019

RESULT 13
US-08-728-323A-1
Sequence 1, Application US/08728323A
Patent No. 5948676
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: Immediate Early Protein From Kaposi's
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728.323A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3489 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

Qy	260	gaggacacacaggtcttcaggagagacatgaagtcagattgaagggtccacgttgggccta	319
Db	20886	GAGGATGACGAGGAGGATGCACGAGGAGGATGACGAGGAGGATGACGAGGAGGAGGACGAG	20827
Qy	320	gaggtcttgagctggatgatgaggagcgtgtgcccaaggagtttccaatcccccgat	379
Db	20826	GAGGAGGACGAGGAGGAGGACGAGGAGGAGGAGGAGGAGGATCAGCATGATGAGGAC	20767

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 7, 2002, 23:30:57 ; Search time 3543.72 Seconds
(without alignments)
7768.866 Million cell updates/sec

Title: US-09-911-667A-1
Perfect score: 2562
Sequence: 1 cccagcgtccgggtttgt.....tgaggactgagngccaagtq 2562

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST: *
1: em_estfun: *
2: em_esthum: *
3: em_estim: *
4: em_estom: *
5: em_estpl: *
6: em_estba: *
7: em_estro: *
8: em_estov: *
9: em_hic: *
10: gb_estl: *
11: gb_est2: *
12: gb_hic: *
13: gb_gss: *
14: em_gss_fun: *
15: em_gss_hum: *
16: em_gss_inv: *
17: em_gss_pin: *
18: em_gss_pro: *
19: em_gss_rtd: *
20: em_gss_vrt: *
21: em_gss_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query %	Length	DB ID	Description
C 1	941.4	36.7	972	10	AL523335	AL523335
C 2	882.2	34.4	1022	10	AL523336	AL523336
C 3	736.6	28.8	778	10	AL529154	AL529154
C 4	599.6	23.4	642	10	AL563357	AL563357
C 5	556.8	21.7	578	10	AL563357	AL563357
C 6	553.2	21.6	638	11	AL663941	AL663941
C 7	534.6	20.9	596	10	AL523339	AL523339
C 8	502	19.6	536	10	AL523339	AL523339
C 9	496.4	19.4	570	11	AL523339	AL523339
C 10	463.6	18.1	487	10	AL523339	AL523339
C 11	423.2	16.5	468	11	AL523339	AL523339
C 12	403.8	15.8	662	11	AL523339	AL523339

13	364.8	14.2	371	10	AA349847	AA349847 EST57001
14	358.2	14.0	366	11	T08271	T08271 EST06162 In
15	323.2	12.6	355	10	A1372624	A1372624 EST175519
16	317	12.4	402	10	AA350545	AA350545 EST57879
C 17	301.8	11.8	450	11	BT296940	BT296940 UI-R-DKO-
C 18	287.6	11.2	308	11	BG056004	BG056004 naf86b06.
C 19	286.8	11.2	377	10	AW345918	AW345918 15969 MAR
C 20	279.8	10.9	420	11	BF412623	BF412623 UI-R-BT1-
C 21	278.4	10.9	941	10	AU067579	AU067579 AU067579
C 22	276.6	10.8	418	10	BE097508	BE097508 UI-R-B01-
C 23	259.4	10.1	266	10	AA351693	AA351693 EST59510
C 24	255.4	10.0	258	11	F08438	F08438 SEC20B081 n
C 25	252.4	9.9	239	11	T10079	T10079 seq1296 b4H
C 26	244.8	9.6	289	10	AW18091	AW18091 55608 MAR
C 27	201.8	7.9	489	11	BG820031	BG820031 602782582
C 28	193.6	7.6	360	10	BE656418	BE656418 UI-M-BHO-
C 29	164.8	6.4	742	10	AU220193	AU220193 AU220193
C 30	163.6	6.4	328	11	BT134483	BT134483 UI-M-BH3-
C 31	142.6	5.6	1417	12	AK003981	AK003981 Mus muscu
C 32	138	5.4	441	11	T08675	T08675 EST06567 In
C 33	129.6	5.1	253	10	BB586972	BB586972 BB586972
C 34	124.6	4.9	686	10	BE590388	BE590388 S8128 Sug
C 35	122.4	4.8	489	13	AZ979001	AZ979001 2M0255J05
C 36	121.2	4.7	275	10	BB593243	BB593243 BB593243
C 37	115.2	4.5	208	10	BB590911	BB590911 BB590911
C 38	104.6	4.1	996	13	CNS010GB	AL167780 Tetraodon
C 39	104.6	4.1	1047	13	CNS051V9	AL343278 Tetraodon
C 40	100.6	3.9	300	11	C57850	C57850 C57850 Yuj1
C 41	97	3.8	223	10	BB591622	BB591622 BB591622
C 42	95	3.7	283	10	BB592855	BB592855 BB592855
C 43	92.8	3.6	868	13	CNS04275	AL313898 Tetraodon
C 44	92.2	3.6	300	10	BB266609	BB266609 BB266609
C 45	89.2	3.5	266	10	BB595082	BB595082 BB595082

ALIGNMENTS

RESULT 1
AL523335/C

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AL523335 972 bp mRNA EST 13-FEB-2001
AL523335 LTI_NFL003_NBC3 Homo sapiens cDNA clone CSODC001YN02 3
Prime, mRNA sequence.

AL523335

AL523335.1 GI:12786828

EST.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

Contact: Genoscope

Genoscope Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

Location/Qualifiers

1..972

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CSODC001YN02"

/clone_lib="LTI_NFL003_NBC3"

/sex="male"

/tissue.type="neuroblastoma cells"

/lab_host="DH10B"

/note="organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA

was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-stranded cDNA was digested with Not I and

cloned into the Not I and Eco RV sites of the pCMVSPORT 6

vector. Library was normalized. Library was constructed

by Life Technologies. Contact : Feng Liang Life

Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 236 a 283 c 252 g 194 t 7 others
ORIGIN

Query Match 36.7%; Score 941.4; DB 10; Length 972;
Best Local Similarity 98.3%; Pred. No. 8.8e-232;
Matches 956; Conservative 4; Mismatches 12; Indels 1; Gaps 1;

Qy 951 tctgctcctgaagtcgaagtgatgtgctgctcagggaaccaggaagcaatcgc 1010
Db 972 TCTGGCTGCCGTGAAGTGAAGGTATGTGTCTGTCTGAGGAACAGGAAAA-GCAATCG 914

Qy 1011 ccaccttaaaagagatagaactaaacaggagctcccatccgctggggaactcatca 1070
Db 913 CCACCTTAAGAGGATAGCAACTGAAACGGGGTCCCATGCCGTGGGAAACTCATCW 854

Qy 1071 tctccagacaggaacagcaggcaaaatgaggacctttcacaccccatatttagatgga 1130
Db 853 TCTCCAGACAGACAGCGAGGCAAAATGAGGACCTTCTCACACCCCATTTTAGATGGA 794

Qy 1131 caactttgctgctggtttatatgtgttttccaatgcattcttactacgggttagttc 1190
Db 793 CAACCTTYGCTGCTGGTTTATATGGTTTTCCAATGCATTCTTACTACGGGTTAGTTC 734

Qy 1191 tactcacacagaactctccagcagagagatgtctcggcatctccagtcggaagaagg 1250
Db 733 TACTCACACAGMACTCTTCAGGACAGAGATGTCTGGGGCTCTCCAGTCCGAAGAAGG 674

Qy 1251 ctgtagagcgaataatgcagctgcctgcgagtagctgagtgagagagattacatggact 1310
Db 673 CTGTAGAGGCAAAATGCAGCTGCCCTGCGAGTACCTCAGTGAGGAGATACATGGACT 614

Qy 1311 tgctgtggaacacccctctgaagttccagtgctcctgtgactctgtggattatggacc 1370
Db 613 TGCTGTGGACCACTCTCTGAGTTTCCAGGTGTCTGTGACTCTGTGGATTTATGACC 554

Qy 1371 gccgggggcgaagaagacacatgccctgtgcttctgcatctctctctgcagctcc 1430
Db 553 GCCTGGGGCGCAAGAACACCATGCCCTGTGCTTTTGTCATCTTCTCTCTGCGACCTCC 494

Qy 1431 tgctgtttatctgtgttggagaataatgtctcactctgttactcttcaattcaagagcgt 1490
Db 493 TGCTGTATTATCTGTGTGGAGAAATGTGCTTACTCTGTTACTCTTCATTCAGAGCGT 434

Qy 1491 ttattcttgaggctttcaagcggcatatgtttacacacctgaggtctaccccaaggcaa 1550
Db 433 TTATTTCTGGAGGCTTTCAAGCGGCATATGTTTACACACCTGAGGTCTACCCACGGCAA 374

Qy 1551 cgcgggcccctgcctgggacacgcagcgagcatggcaagagtggtgctctcatcactc 1610
Db 373 CGCGGGCCCTCGGCTGGGACCTTCGAGCGCATGGCAAGAGTGGGTGCTCTCANTACTC 314

Qy 1611 cgttcatgccaggtgatcgtggaatcctctgtgtacctgactctgtgacgtttacagtg 1670
Db 313 CGTTATCGCCAGGTGATGCTGGAATCCTCTGTGTACTCTGACCTCTGGCAGTTTACAGTG 254

Qy 1671 gctgctcctctgctgctcctgctcctgctcttttttggccattgagaccaaagcggag 1730
Db 253 GCTGCTGCCTCCTGGCTGCCCTGGCTCCTCTGCTTTTGGCCCATTTGAGACCAAGGCCGAG 194

Qy 1731 gactgcaggagatccagccacacgggagtgggccaggagatggtcggccgaggaatgcacg 1790
Db 193 CACTGCAGGAGTCCAGCCACCGGAGTGGGGCCAGGAGATGTTGCGCCGAGGATGCACG 134

Qy 1791 gtgcaggtgttacacaggtcgaactctgtgctctcaaggaatagtcaccagatggggactgag 1850
Db 133 GTGCAGGTGTTACAGGTGCAACTCTGGCTCTCAGGAATAGTGACCATGGGGACTGAG 74

Qy 1851 ctggctctttgaggctgcagagcttgggggctggtgcagggcccaactggggcactgattg 1910

Db 73 CTGGTCTTTGAGCTGCAGAGCTTGGGGGCTGGGAGCCCGACCTGGGGCACTGATTGT 14

Qy 1911 cactgcccacatc 1923
Db 13 TACTCCCGACATC 1

RESULT 2
AL523336 1022 bp mRNA EST 13-FEB-2001
LOCUS AL523336 LTI_NFL003_NBC3 Homo sapiens cDNA clone CS0DC001YN02 5
DEFINITION prime, mRNA sequence.
ACCESSION AL523336
VERSION AL523336.1 GI:12786829
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1022)
AUTHORS Li, W. B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..1022
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DC001YN02"
/clone_lib="LTI_NFL003_NBC3"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 180 a 268 c 326 g 243 t 5 others
ORIGIN

Query Match 34.4%; Score 882.2; DB 10; Length 1022;
Best Local Similarity 98.0%; Pred. No. 1.7e-216;
Matches 952; Conservative 4; Mismatches 7; Indels 8; Gaps 6;

Qy 13 ggccttctcgtggctgggtcccccagcgctccctcccccaaacagctgctccagg 72
Db 59 GCCTTTGTCTGTGGCTGGTCCCGCAGCGGTCTCCCTCCCGAACAGCTGCTCCAGGG 118

Qy 73 aggaagcgcgcgg-gtgcctcagacttcccgtgtctgaaacggagggtcgtcatc 131
Db 119 AGGAAGCGGCGGNGTCTGTCCAGCTTCCCGGTGCTGAAACCGGAGGGCTCGTCATC 178

Qy 132 caccactaccatgtaagggcccatgagaagggtcctatcctgcgcagcgcgcagacatggagg 191
Db 179 CACCACATACCATGTAAAGGCCCATGAGAGGGCTCATCTCTGCGCA-SGCGACATGGAGG 237

Qy 192 aggacttattccagtaaggcagctgcggttgtgaaattccgtgcacagcgagagtg 251
Db 238 AGGACTTATTCCAGCTAAGGACGTGCCGGTGTGAAATTCGTCGCACAGCGAGAGTG 297

Qy 252 caaggtcagagacacacgcttcagagagcatgaaagtccagattgaaggggtccacg 311


```

QY 732 tgggttcgggagtcgagagtgctcccgctcggtgacgtgtatgccga 780
|||||
Db 730 TGGGCTTCGGGATCGGAGGAGTTCGCCAGTTCGGTGACGCTTKTTGCCGA 778

RESULT 4
AL563357 642 bp mRNA EST 16-FEB-2001
LOCUS AL563357.LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DD004YD17 3
DEFINITION prime, mRNA sequence.
ACCESSION AL563357
VERSION AL563357.1 GI:12912671
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 642)
AUTHORS Li.W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
Location/Qualifiers
1..642
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DD004YD17"
/clone_lib="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 149 a 192 c 170 g 117 t 14 others
ORIGIN

Query Match 23.4%; Score 599.6; DB 10; Length 642;
Best Local Similarity 96.0%; Pred. No. 9.9e-144;
Matches 616; Conservative 2; Mismatches 23; Indels 1; Gaps 1;

QY 1221 atgtctcgccatctccagtcgaagaggtctgttagaggcaaaatgcagctggcctgcg 1280
|||||
Db 642 ATTTCTCGGGCTTCTCCAGTCGGAARAAGGCTGAAGAGGCAAAATGCAGCTGGCTGCG 583

QY 1281 aatacctgagtgagagattacatgagctgtgtgaccacccctctctagttccag 1340
|||||
Db 582 ACTACTGAGTGAGGAGATWACATGGACTTGTCTGTGACCACCCCTCTCTGAGTTTCCAG 523

QY 1341 gtgtcctgtgactctgtgattattgaccgctggcgcaagaagaccatggccctgt 1400
|||||
Db 522 GTGTCTTGACTGTGTGATTATTGACCGCTGGGGCGCAAGAACCATGGCCCTGG 463

QY 1401 gcttgcattctctctctgcagcctctgtgtttattctgtgttggaag-aaatgtg 1459
|||||
Db 462 GCTTTGTCTATCTCTCTTCGACGCCNCCNGCTGTTNNCTGTGTGGAAGAAATGNG 403

QY 1460 ctactctgttactcttcattgcaagacgtttattcttgaggctttcaagcgcatat 1519
|||||
Db 402 CTCACCTCTGTAACTCTTCATGGCAAGACGCTTATTTCTGAGGCTTTTCAAGCGGCATAT 343

QY 1520 gtttacacacctgaggtgtactaccacgcaagcgcgccctcgccctgggcacctgcagc 1579
|||||

```

```

Db 342 GTTNACACACCTCAGGTCTACCCACGCAACGCGGCCCTCGCTGGCACCTGCAGC 283
|||||
QY 1580 ggcattgcaagagtggtgtctctcatcactccgttcatccagagtgatgctggaatcc 1639
|||||
Db 282 GGCATGGCAAGAGNGGGGCTCTCANTACTCCGTTTCATGCCCCAGGTGATGCTGGAATCC 223
|||||
QY 1640 tctgtgacctgactgacgtttacagtggctgctgctcctgctgctgctccctggcctcc 1699
|||||
Db 222 TCTGTGTACTCTGACTCTGGCAGTTAACAGTGGCNGCTGCTCTGCTGCCCTGGCCTCC 163
|||||
QY 1700 tgccttttgcctattgagaccaaagcgagagactcagagagtcacagccacccggagtg 1759
|||||
Db 162 TGCCTTTTGCCTCATTGAGACCAAGCCGAGNACTCCAGAGTCCAGCCACCGGAGTGG 103
|||||
QY 1760 gcccaggagatggtcgccgaggaatgcacggtgcaggtgtttaccaggtgcgaactctg 1819
|||||
Db 102 GGCCAGGAGATGGTGGCCGAGGAATGCACGCTGCAAGGTGTTACCAAGTGGACTCTGGC 43
|||||
QY 1820 tctcaggaatagtgaccgatggggactgagctgagctggtctttga 1861
|||||
Db 42 TCTCAGGAATAGTACCGATGGGGGACTGNGCTGGTCTTTGA 1

RESULT 5
AL5633941/c
LOCUS AW663941 578 bp mRNA EST 06-APR-2000
DEFINITION hi73h10.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2977987 3' similar to TR:Q9Z2I7 Q9Z2I7 SV2 RELATED PROTEIN.
; mRNA sequence.
ACCESSION AW663941
VERSION AW663941.1 GI:7456480
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 578)
REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 464.
Location/Qualifiers
1..578
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2977987"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pF73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI-CGAP-GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 137 a 170 c 159 g 111 t 1 others
ORIGIN

```

Query Match 21.7%; Score 556.8; DB 10; Length 578;
Best Local Similarity 99.3%; Pred. No. 1e-132;


```
VERSION      AW957713.1  GI:8147396
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
REFERENCE     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
              1 (bases 1 to 596)
AUTHORS      Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
              ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
              Quackenbush,J.
TITLE        Assessment of gene expression patterns in a model of colon tumor
              metastasis using a 19,200 element cDNA microarray
JOURNAL
COMMENT      Contact: John Quackenbush
              The Institute for Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850, USA
              Tel: 301 838 3528
              Fax: 301 838 0208
              Email: johnq@tigr.org
              Plate: 110
              Seq primer: Reverse.
              Location/Qualifiers
              Source
                1..596
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_lib="MAGE resequences, MAGE"
                /note="vector: pBluescriptskm"
BASE COUNT   133 a 170 c 149 g 144 t
ORIGIN
Query Match  20.9%; Score 534.6; DB 10; Length 596;
Best Local Similarity 97.6%; Pred. No. 5.5e-127;
Matches 574; Conservative 0; Mismatches 10; Indels 4; Gaps 3;
QY 1883 ggcagggcccaactggggcactgattgtcactgcgcacatcaagaactcaccagaagta 1942
DB 11 GGCAGGCCCAACTGGGGCACTGATTGTCACTGCGGACATCAAGAACTCACCACAGAGTA 70
QY 1943 tgacctggaccacacagggtttgtgtcttgactcagttgtgtcatcttcattgaggtcca 2002
DB 71 TGACCTGGACCAACAGGGGTTTGTGTCTTGACTCAGTTTGTCTCATCTTCATTGAGGTCCA 130
QY 2003 cccagggatggagatgtttgtctctagggggttctctgtatatgtgtgaaactttgt 2062
DB 131 CCCAGGGATGGGAGATGTTTGTCTCTAGGGGTTCTCTGTATATGTGTGTGAACCTTTGT 190
QY 2063 tcataacctgtggtatctacatgggaagactacccalattaggagggtctgtaatgccag 2122
DB 191 TCATAAACCTGTGGATCTACATGGGAAGACTACCCATATTAGGAGGGTCTGGTAATGCCAG 250
QY 2123 caaccaatcacaccacacagatcacccgccaacccctcaatgaacacccaataat 2182
DB 251 CAACCAATCAGACACACCCAGAGTACCCGGCCAAACCCCTCAGTGAACACCAAAATAT 310
QY 2183 ctctctgtagataccgtccaggctcaggcccatgtgacacctgtgtccaccaccaggac 2242
DB 311 CTCCTGTAGATACGTCAGGCTCAGGCCCATGTGTGACACCTGTCTCCACCCACCGGAC 370
QY 2243 ctgttcagtaggtttctccacacacacacagcccccaggtttctctttgaaatgcagcc 2302
DB 371 CTGTTTCAGTAGGTTTCTCCCAACACCCACAGGCCCCAGGGTCTTCTTTTGAATTCAGGC 430
QY 2303 gatctagggtggttctgagcagctattctctggcaggggccccccgggtttgctccctta 2361
DB 431 GATCTAGGTGTGGTCTTTCAGCAGCTATTTCCTGGCAGGGGCCCGCGGTTCCTTCTTA 490
QY 2362 gagcctgaccagtgattctcttgccagatggacatggtgca-ttcaaaactggagccaat 2420
DB 491 GAGCCTGACCATGATTTCTCTGGCAGATGGACATGGTGTGCAATTTCAAACACTGGAGCCACAT 550
QY 2421 qcscacacacacccctnttgaggttgccgtttgttgccacccaagaga 2468
|||||
```

```
Db 551 GCCCCCACCCATCCCCCTT--GGAGTGTTCGTTGCTTGTGGCACCACGAAGA 596
RESULT      8
AW965645    536 bp  mRNA  EST  01-JUN-2000
LOCUS       EST377718 MAGE resequences, MAGI Homo sapiens cDNA, mRNA sequence.
DEFINITION  AW965645
ACCESSION   AW965645
VERSION     AW965645.1  GI:8155481
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
              1 (bases 1 to 536)
REFERENCE   1
AUTHORS     Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
              ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
              Quackenbush,J.
TITLE        Assessment of gene expression patterns in a model of colon tumor
              metastasis using a 19,200 element cDNA microarray
JOURNAL
COMMENT      Contact: John Quackenbush
              The Institute for Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850, USA
              Tel: 301 838 3528
              Fax: 301 838 0208
              Email: johnq@tigr.org
              Plate: 217
              Seq primer: Reverse.
              Location/Qualifiers
              Source
                1..536
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_lib="MAGE resequences, MAGI"
                /note="vector: pBluescriptskm"
BASE COUNT   95 a 144 c 174 g 123 t
ORIGIN
Query Match  19.6%; Score 502; DB 10; Length 536;
Best Local Similarity 99.0%; Pred. No. 1.3e-118;
Matches 505; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1547 gcaacgcgggcccctcgccctgggcacctgcagcgcatggcaagagtggtgctctc 1606
DB 27 GCCACGGGGGCCCTCGGCGCTGGGCACCTGCAGCGGCATGGCAAGAGTGGGTGCTCTCATC 86
QY 1607 actccgttcatcgcccagggtgatgctggaatctctgtgtacctgactctgacggtttac 1666
DB 87 ACTCCGTTTCATCGGCCAGGTGATGCTGGAATCCTCTGTGTACCTGACTCTGCGAGTTTAC 146
QY 1667 agtgctgctgctcctgcctgcctgcctgcctcctcctgttttggccattgagaccaaagc 1726
DB 147 AGTGGCTGCTCCTCGCTGGCTGCCCTGCGCTTCTGCTCTTTTGGCCATTGACACCAAGGC 206
QY 1727 ggaggactgcaggagtcagccacccgggagtgggggccaggagatggtcgccgaggaatg 1786
DB 207 CGAGGACTGCAGGAGTCCAGCCACCGGAGTGGGGCCAGGAGATGTCGCGCGAGGAATG 266
QY 1787 caccgtgcaggtgttacccagtgaaactctggtctcctcaggaatgtaaccatggggac 1846
DB 267 CACGGTGCAGGTGTTCACAGGTTCGAACCTGTGGCTCTCAGGAATAGTACCCATGGGGGAC 326
QY 1847 tgagctggtctttgaggtcgagagcttggggggtggcgagggcccaactggggcactga 1906
DB 327 TGAGCTGGTCTTTGAGGCTGCAGAGCTTGGGGGCTGCGAGGCCCACTGGGGCACTGA 386
QY 1907 ttgtcactgcgcacatcaagaactcaccacagagatgatgacctgagccaaacgggtttgt 1966
DB 387 TTGTCACTGCGGACATCAAGAACTCACCAGAGTATGACCTGGACCAACAGGGTTTGT 446
QY 1967 gctc-tactcaotttgcattctcattgaggttcaccacaggatggggagatggttct 2026
|||||
```


Query Match 18.1%; Score 463.6; DB 10; Length 487;
Best Local Similarity 96.9%; Pred. No. 1e-108;
Matches 472; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1340 ggtatccttgactctgtgattattgacgcctgagcgccgagcgaagacatggccctg 1399
Db 487 GGTGTGCTATGACTCTGTGATTTATGACCGATGGGGGCAAGACCATGGCCNTG 428
Qy 1400 tgccttgatctctctctctcgcagcctcctcgtgttattctgtgtggaagaaatgtg 1459
Db 427 TGTGTGCTATGACTCTGTGATTTATGACCGATGGGGGCAAGACCATGGCCNTG 368
Qy 1460 ctacactgttactctctcctcgaagagcgtttattctgagagctttcaacgagcatat 1519
Db 367 CTCACCTGTGTTACTCTTCAATTCGAAGAGCGTATTATGAGGAGCTTTCAACCGGCATAT 308
Qy 1520 gtttacacacctgaggtctacccacgcaacgagcgccctcgcctgggacacctgcagc 1579
Db 307 GTTTACACACGTGAGGTCTACCCACGCAACGCGGGCCCTCGGCCCTGGGCACCTGCAGC 248
Qy 1580 ggcattgcaagagtggtgctctctcatcactcctgcttcacgcccaggtgagctgaaatcc 1639
Db 247 GGCATGGCAAGAGTGGTGTCTCTATCATCTCCGTTCATCGCCCAAGGTGATGCTGGAATCC 188
Qy 1640 tctgtgacctgactcgtgagcgtttacagtggtcgtcctcctcgtgctgacctgacctcc 1699
Db 187 TCTGTGTAACCTGACTCTGGCAGTTTACAGTGGCTGCTGCTCTCTGGCTGGCTGGCTGCC 128
Qy 1700 tgccttttgcacattgagacaaaggcgaggaactgcagggagtcacgcccagcgagtggtg 1759
Db 127 TGCCTTTTGCCCATTTGAGACCAAGAGCGGAGGACTGCGAGGAGTCCAGCACCGGAGTGG 68
Qy 1760 ggcagagagtggtcgcgcgaggaatgcacggtgcaggtgttaccaggtgcgaactctgac 1819
Db 67 GGCAGGAGAGTGGTGGCCAAAGAAATGACGGTTCAGGTGTCCAGGTTCGCAACTCTGGC 8
Qy 1820 tctcagg 1826
Db 7 TCTCAGG 1

RESULT 11
BF775076 468 bp mRNA EST 25-APR-2001
LOCUS 285128 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
DEFINITION BF775076
ACCESSION BF775076
VERSION BF775076.1 GI:12122976
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 468)
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Perlea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keeler,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers

FEATURES
source
1. .468
Location/Qualifiers
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 3BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."
BASE COUNT 87 a 126 c 139 g 116 t
ORIGIN

Query Match 16.5%; Score 423.2; DB 11; Length 468;
Best Local Similarity 94.0%; Pred. No. 2.6e-98;
Matches 440; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Qy 671 tatggcatccttagtgatttgcccgctgtatagctgagatcctggtcctccgggacctg 730
Db 1 TACGGCATCTCTAGTGGCTTTGGCGCCGTGTACAGCTGGATCCTGGTGTCTCCGGGGCTG 60
Qy 731 gtggcttcgggattcgaggaggttccccagtcggtgacgctgtatgcgaggttctctccc 790
Db 61 GTGGCTTCGGGATCGGAGGGTCCCCACGTCGGTCACACTGTATGCTGAGTTCCTCTCC 120
Qy 791 atgaagccagagctaaatattttctgctgattgaggtattcttgccatcgagacagt 850
Db 121 ATGAAGCCAGAGCTAAATGTTTGTGATGATGAGGTGTTCTGGGCAATCGGACAGTG 180
Qy 851 ttcagatcctcgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctc 910
Db 181 TTCAGGTGCTGCTGGCTGTGTTTGTGATGCCAGCTGGCTGGCTGGCTGGCTGGCT 240
Qy 911 ctctcagctgtcccnctcctcctcttctgctgctgctgctgctgctgctgctgctgctg 970
Db 241 CTCACAGTGTCCACCTCTCTCTTTGGCTGCTCTGTTTGGCTGCCGAGAGTGGC 300
Qy 971 aggtatgctgctgctgaggaacaggaagaaacaggaagaaacaggaagaaacaggaag 1030
Db 301 AGGTATGATGCTGTCCGGAAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360
Qy 1031 actgaaacagagctcccatgctgctgctgctgctgctgctgctgctgctgctgctgct 1090
Db 361 ACGGAAACAGGAGCTCCATGCCCTCTGGGAAACTCATCATTTCCAGACAGGAGACGA 420
Qy 1091 ggcataatgagggaccttttccacacccatttttagatggagaaactttg 1138
Db 421 GCGAAATAGGGAGCTTTTTCACACCCCATTTTATAGTAGGACAACTTG 468

RESULT 12
BF775076 662 bp mRNA EST 06-AUG-2001
LOCUS 285128 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
DEFINITION BF775076
ACCESSION BF775076
VERSION BF775076.1 GI:15087284
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 662)
Porter,T.E. and Cogburn,L.A.

QY	1959	gggtttt 1964
Db	361	GGTTT 366
RESULT	15	
AI372624		
LOCUS	EST175519 Infant brain, Bento Soares Homo sapiens cDNA clone	12-JAN-1999
DEFINITION	UHIBZ2R 5' end, mRNA sequence.	
ACCESSION	AI372624	
VERSION	AI372624.1 GI:4152490	
KEYWORDS	EST.	
SOURCE	Homo sapiens	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 355)	
AUTHORS	Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult .C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White .O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald .L.M., Fitzhugh,W.M., Frichman,J.L., Geoghagen,N.S., Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaro,S.M., Merrick,J.M., Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y., Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W., .Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,H.K., Kozak,D.L., Kunsch,C., HungJun,J., Li,H., Weissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillilion,P.J., Fannon .M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.	
TITLE	Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence	
JOURNAL	Nature 377 (6547 Suppl), 3-174 (1995)	
MEDLINE	96026280	
COMMENT	Other ESTs: THC98959 Contact: HGI (Human Gene Index) The Institute for Genomic Research 9712, Medical Center Drive, Rockville, MD 20850, USA Tel: (301)-838-0200 Fax: (301)-838-0208 Email: hgi@tigr.org Seq primer: M13 Reverse. Location/Qualifiers 1..355 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="UHIBZ2R" /lab_host="E. coli DH5-alpha" /note="Vector: BA, M13-derived; Site_1: HindIII; Site_2: NotI; The infant brain library, constructed by Bento Soares, Columbia University, was oligo-(dT) primed and directionally cloned into an M13-derived plasmid using total brain mRNA from a 72-day old human female afflicted with spinal muscular atrophy."	
FEATURES	source	
BASE COUNT	66 a 93 c 111 g 75 t	10 others
ORIGIN		
Query Match	12.6%	Score 323.2; DB 10; Length 355;
Best Local Similarity	96.0%	Pred. No. 1.4e-72;
Matches	336; Conservative	0; Mismatches 13; Indels 1; Gaps 1
Qy	1594	gggtgctctcatcactcgcttcacgccccaggtagctgcagatcctctgtgtacctgac 1653
Db	7	GAGGGCTTCATCATCTCCGTTCATCGGCCCAAGGTAGCTCGGAATCCCTCTGTGTACCTGCAC 66
Ov	1654	tctgcagatttaccagtgcctgcctgcctcctgcctgcctgcctgcctgcctgcctgttttgcccat 1713

Search completed: March 8, 2002, 00:47:58
Job time: 4621 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 8, 2002, 03:28:23 ; Search time 5957.55 Seconds
(without alignments)
4552.436 Million cell updates/sec

Title: us-09-911-667A-3
Perfect score: 1644
Sequence: 1 atggaggaggaactattcca.....cgaactctggctctcaggaa 1644

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_om.*
- 20: em_or.*
- 21: em_ov.*
- 22: em_pat.*
- 23: em_ph.*
- 24: em_pl.*
- 25: em_ro.*
- 26: em_sts.*
- 27: em_sy.*
- 28: em_un.*
- 29: em_vl.*
- 30: em_htgo_hum.*
- 31: em_htgo_inv.*
- 32: em_htgo_rod.*
- 33: em_htg_hum.*
- 34: em_htg_inv.*
- 35: em_htg_rod.*
- 36: em_htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB	ID	Description	
1	1326.2	80.7	2421	10	AF060173	AF060173 Rattus no	
2	432.4	26.3	2104	9	HS0802697	AL359592 Homo sapi	
3	209.4	12.7	212183	2	AC087893	AC087893 Homo sapi	
c	91	5.5	11856	1	AE007266	AE007266 Sinorhizo	
	89.4	5.4	148432	3	AC004642	AC004642 Drosophill	
	89.4	5.4	157851	2	AC020509	AC020509 Drosophill	
6	89.4	5.4	300542	3	AE003462	AE003462 Drosophill	
c	7	89.4	5.4	198677	1	AE001863	AE001863 Deinococc
8	89	5.1	19834	3	CEK01F9	222175 Caenorhabdi	
c	9	84	5.1	10344	1	AE004767	AE004767 Pseudomon
10	82.8	5.0	10344	2	AC067757	AC067757 Homo sapi	
c	11	81.8	5.0	62204	2	AE008141	AE008141 Agrobacte
12	77	4.7	11286	1	AB026645	AB026645 Arabidops	
13	74.2	4.5	81020	8	AB026645	AB026645 Arabidops	
14	74.2	4.5	81020	24	AC024128	AC024128 Arabidops	
15	70	4.3	13462	1	AF218267	AF218267 Pseudomon	
16	69.6	4.2	2622	10	AF060174	AF060174 Rattus no	
17	69.2	4.2	1638	9	AF052188	AF052188 Homo sapi	
18	69.2	4.2	4948	9	AB018278	AB018278 Homo sapi	
19	66.2	4.0	2315	5	DYGSV2TRAN	L23403 Discopysge o	
20	66.2	4.0	3190	9	BC000776	BC000776 Homo sapi	
21	66.2	4.0	4353	9	AB018279	AB018279 Homo sapi	
22	66	4.0	3660	10	RATSV2BA	L10362 Rattus norv	
23	64	3.9	35000	1	AB000617	AB000617 Bacillus	
24	64	3.9	221160	1	BSUB0002	299105 Bacillus su	
25	63	3.8	2268	6	AX079057	AX079057 Sequence	
26	62.4	3.8	10029	1	AE007908	AE007908 Agrobacte	
27	61.6	3.7	16775	1	AB0000735	AB0000735 Nocardiol	
c	28	61.6	3.7	40699	3	CEZK637	211115 Caenorhabdi
29	59.8	3.6	7424	1	AB050935	AB050935 Pseudomon	
c	30	59.2	3.6	11886	1	AE004638	AE004638 Pseudomon
31	58	3.5	159167	2	DMBR42L18	AL122031 Drosophill	
32	57.2	3.5	37221	2	AC012886	AC012886 Drosophill	
33	57.2	3.5	126340	2	DMBR18C8	AL122030 Drosophill	
c	34	57.2	3.5	299537	3	AE003438	AE003438 Drosophill
35	56.6	3.4	3868	4	S47919	AF030325 Comamonas	
36	56.6	3.4	7080	1	AF305325	AF305325 Streptomyc	
37	56.4	3.4	35654	1	SC7H1	AL021411 Thermotog	
38	53.8	3.3	17641	1	AE001804	AE001804 Sequence 5	
39	53.2	3.2	1885	6	A48158	A48158 Sequence 5	
40	51	3.1	134092	2	H0702G05	AL442105 Oryza sat	
41	50.8	3.1	3663	10	RATSV2A	L01788 Rattus norv	
42	50.8	3.1	3844	10	RATSV2AA	L05435 Rattus norv	
43	50.2	3.1	2132	10	RNOCTK	X98334 R.norvegicu	
c	44	49.8	3.0	12460	1	AE004903	AE004903 Pseudomonas
45	49.2	3.0	3116	1	PPU10895	U10895 Pseudomonas	

ALIGNMENTS

RESULT	1				
AF060173	AF060173	2421 bp	mrna	ROD	21-NOV-1998
LOCUS	Rattus norvegicus	SV2 related protein (SVOP)	complete cds.		
DEFINITION	Rattus norvegicus				
ACCESSION	AF060173				
VERSION	AF060173.1	GI:3901267			
KEYWORDS	Norway rat.				
SOURCE	Rattus norvegicus				
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;				
	Rattus.				
REFERENCE	1 (bases 1 to 2421)				
AUTHORS	Janz,R., Hofmann,K. and Sudhof,T.C.				
TITLE	SVOP, an evolutionarily conserved synaptic vesicle protein,				
	suggests novel transport functions of synaptic vesicles				
JOURNAL	J. Neurosci. 18 (22), 9269-9281 (1998)				
MEDLINE	990197;5				
REFERENCE	2 (bases 1 to 2421)				
AUTHORS	Janz,R., Hofmann,K. and Sudhof,T.C.				

TITLE Direct Submission
JOURNAL Submitted (16-APR-1998) Center for Basic Neuroscience, Department of Molecular Genetics, HHMI, UT Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, TX 75235, USA

FEATURES
source location/Qualifiers
1..2421
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/tissue_type="brain"
1..2421
/gene="SVOP"
301..1947
/gene="SVOP"
/note="synaptic vesicle protein; similar to transport proteins"
/codon_start=1
/product="SV2 related protein"
/protein_id="AAC78627.1"
/db_xref="GI:3901268"
/translation="MEEDLFQRLPVPVFRRTGESARSEDDAASGEHDVQIEGVRVG LEAVELDGAAPKPEFANPTDTEFMVEDAVEAIGFGRFQWKLISVLTGLAMMADAMEMH ILSILAPQLHCHEMLPSQVALLTSVVFIGNMSSITLGNISDQYGRKTLGISVFWT LYGILSAFAHPVSMILVLRGLVGFVGIVPOSVTLVAEFLPMKARAKCILLIEFWA IGTVEVLLAVFVPSLSGRWLLLSAAILVFAVLCFWLPESARYDLVLSQNEKATA TLKRIATENGAPMLGLIISKQEDRGKMRDLTPHPRWTFLLWFLWFSNAFSYGL VLLTELFQAGDVCSI SRKKAVERAKSLACEYLSKEDYMDLLWTLSEFPQVLTLM VIDRLGRKTMALCFVIFSLCSILLIFICIGRNVLTLLLFTARAFISGGFOAYVYTFE VYPTATRALGLGTCSMARVGLITPPIAQWLESVYLLTAVYSGCLLAAALASCLF PIETKRALQBSHREMQEVMGRGTNSVTPRNSGSQE"

BASE COUNT 459 a 658 c 735 g 569 t

ORIGIN

Query Match 80.7%; Score 1326.2; DB 10; Length 2421;
Best Local Similarity 87.9%; Pred. No. 0;
Matches 1445; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

Qy 1 atggaggaggacttattccagtaaggcaagctgcgggtgtgaaattccgtgcacagc 60
Db ATGGAGGAGGACTTGTTCAGCTCAGGCAGTTGCCGGTGTGTAATTCGGCGCACAGGA 360
Qy 61 gagagtgcgaagtcagagacacacgcttcagagagacatgaagtcacagattgaagg 120
Db GAGAGCGCAGCTCAGAGACACGCGCTTCCGGGGAACATGATGTTCCAGATTGAGGG 420
Qy 121 gtccagtgaggcctgagagctgagagctgagatgagtgaggcagctgccccaaagagttl 180
Db GTCCGAGTGGGCTAGAAAGCTCAGAGCTGGATGGAGCAGCTGTCCCCCAAGGAGTTT 480
Qy 181 gccaatcccaacagatgatacttccatggtggaagatgcagtggaagccattggcttggg 240
Db GCCAATCCCACTGATGACACTTTTCATGGTCAAGATGCGCGTGGAGGCCATTGGGTTCGA 540
Qy 241 aaatttcagtggaagctgctgtctctacatggcttggttgatg,jctaatgccatggag 300
Db AGATTCCAGTGAAGCTCTCTCTCTCAGCGGCTTGGCTTGGATGGCGGACACGTCATGGAG 600
Qy 301 atgatgattcctcagatcctggcacacacagctgcatctgcagatggaggtcccaagctg 360
Db ATGATGATTCTCAGGATCTCGGCGCTCAGCTGCATCTGCAGTGGCGGACTCCCGACTGG 660
Qy 361 caggtyggcattgtagctcgtggtcttcttgtaagcatgatgccagctccacgctctgg 420
Db CAGGTGGCGCTGCTGACTTCGGTGGTCTTCATTTGATGATGATGATGATGATGATGATGAT 720
Qy 421 gaaatatctcagacagctacgcaggaagaaacagggtcgaagatcagctgctgagact 480
Db GAAACATCTCGGATCAGTATGGCAGGAAACAGGGCTGAAGATCAGTGTGTTCTGGAGCC 780
Qy 481 ctgtactatggcattccttagtcttgcgcccgtgtagctgagatcctgtagctccgg 540
Db CTGTACTACGGCATCCTCAGCGCTTTCGCGCCAGTGTATAGCTGGATCTGTGTGTCCTCCA 840

Qy 541 ggcctggtggtggtccttcggagtcggagaggttccccaagtcggtagcagctgtatgccagttc 600
Db GGCCTCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 900
Qy 601 ctcccatgaaagccagagactaaatgtatttgcgtgattgaggtattctgggcccattcg 660
Db CTCCCATGAAGCCAGAGCCCAAGTGCATTTTCTGATTGAGGTTTCTGGGCCATCGGG 960
Qy 661 acagtgttcgaggtcgtcctggtgctgtgttcgtagccacgctgggtggcgttggtg 720
Db ACCGTGTTTCGAGGTGCTTCTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1020
Qy 721 ctacatcctctcagctgctccgctcctctctctctctctctctctctctctctctctctgaa 780
Db CTGCTGCTGTCTGGCGCGCTCCACTACTTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 1080
Qy 781 agtcaaggtatgctgctgcagggaaccaggaagaaagcaatcgccaccttaagagg 840
Db AGTGCTAGGTAGTGTGCTGTCTGGGAACACAGAAAGGCCATTCCTAAAGCGG 1140
Qy 841 atagcaactgaaacggagctccccatgcgctggggaactcatcatctccagacagga 900
Db ATCGCAACAGAAATGGAGCCCATGCTCTCTGGGAAGCTCATCATCTCCAGACAGSAA 1200
Qy 901 aaccaggaacaaatgagagacctttcacaccccatitttagatgagacaactttgctgctg 960
Db GACCGAGGCAAAATGAGGAGCCCTTTTACACCCCACTTTCTGTTGGACAACACTCTGCTGCTG 1260
Qy 961 tggtttatatggttttcccaatgcattctctactacggtttagttctactcaccacagaa 1020
Db TGGTTTATATGTTCTTCCAATGCTTCTCTTATACGGGTGCTTCTGTCTGTCTGTCTGTCTGT 1320
Qy 1021 ctcttctcggcaggagatgctgcggcctcctcagtcgggaagagctgtagagcaaaa 1080
Db CTCCTTCCAGCGCGAGATGTTTGCAGCATCTCCAGCCGGAAGAGCGGTGGAGGCAAA 1380
Qy 1081 tgacgcttggtcgtgagtaacctgagtgagagattacatggaacttgcgtgagaccacc 1140
Db TCGAGCTTGGCTGTGAGTACCTCAGCAAGAGGATTACATGGACCTGCTGTGTGGACCACT 1440
Qy 1141 ctctctgagtttccaggtgctctgtgactctgtggaattattgacgctggtggcgcaag 1200
Db CTGCTGAGTTCCAGGTGCTCTGTGACTCTGTGGGTTCATCGACCGCTTGGCGCCGCAAG 1500
Qy 1201 aagacatgcccctgctgttctgctcctctctcctcctcctcctcctcctcctcctcctcctg 1260
Db AAGACCATGGCTGTGTTTCTGCTCATCTTCTCTCTCTGAGCCCTCTCTGTTCTGTTCTGCTG 1560
Qy 1261 gttggaagaaatgctcactcctgttactcttcatttgcgaagagcgtttatttctgagggc 1320
Db ATTGGAAGAAATGTGTAACCTCTTACTTGTTCATTGCAAGAGCGTTTATTTCTGGAGGC 1620
Qy 1321 ttccaagcggcctatgtttacacacctgaggtctacccacagcggcaagcggcgccctcgagc 1380
Db TTCCAAGAGCGCTACGTTTACAGCGCTGAGGTGTATCCAACGCGCAGACGGGCGCTGGGC 1680
Qy 1381 ctgggcacctcgcagcggcatggcaagtggtgtgctctcctcctcctcctcctcctcctcctg 1440
Db CTGGGCACCTGCAGCGCATGGCAGAGTGGCGCGCTCATCTCATCTCATCTCATCTCATCTCAG 1740
Qy 1441 gtgagtctggaatcctcctgtgtacctgactcgtggcagtttacagtggctgctgctcctg 1500
Db GTGATGCTGGAAATCTTCCGCTTACCTGACCTGGCGCTGTACAGTGGCGTGTGCTGCTGCTT 1800
Qy 1501 gctgcccctgctcctcctgttttggccattgagacaaagcggagagcgtgcaggagctcc 1560
Db GCTGCCCTTGGCCTCTCTGCTTCTGCCCCATCGAGACCAAGGCCGAGCACTGCGAGGAGTCC 1860
Qy 1561 agccaccgggagtgggggccagagatggttcggccgaggaatgcacggtgcaggtgttaacc 1620
Db AGCCACGGGAGTGGGGCCAGGAGATGTTGGCCGAGGAGCAAAACAGACAGCGCGTCCCC 1920
Qy 1621 aggtcgaaactctggctctcagga 1643

Db 1921 AGGTGCAACTCTGGCTCTCAGGA 1943
 |||||||
 RESULT 2
 HSM802697 2104 bp mrna PRI 18-JUN-2000
 LOCUS Homo sapiens mRNA; cDNA DKFZp761H039 (from clone DKFZp761H039).
 ACCESSION AL359592
 VERSION AL359592.1 GI:8655658
 KEYWORDS human.
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 2104)
 AUTHORS Blum,H., Buaersachs,S., Mewes,H.W., Weil,B. and Wiemann,S.
 TITLE Direct Submission
 JOURNAL Submitted (15-JUN-2000) MIPS, Am Klopferspitz 18a, D-82152
 Martinsried, GERMANY
 COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by LMU (Ludwig Maximilians University,
 Munich/Germany) within the cDNA sequencing consortium of the German
 Genome Project.
 This clone (DKFZp761H039) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
 information about the clone and the sequencing project is available
 at http://www.mips.biochem.mpg.de/proj/cDNA/.
 FEATURES
 Location/Qualifiers
 1..2104
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="DKFZp761H039"
 /clone_lib="761 (synonym: hamy2). Vector pSport1; host
 DH10B; sites NotI + SalI"
 /dev_stage="adult"
 /tissue_type="amygdala"
 1..437
 /gene="DKFZp761H039"
 <1..437
 /gene="DKFZp761H039"
 /note="Similarity to SV2 related protein (Rattus
 norvegicus)"
 /codon_start=3
 /product="hypothetical protein"
 /protein_id="CA894878.1"
 /db_xref="GI:8655659"
 /translation="LCFVIFSPCSILLFCVGRNVLTLLIFTARAFISGFORAYVYT
 PVPYPTKRALGLGTCGMARVAGLITPTIAQVMLESSVLTLLAVSGCCLLAALASL
 FPIETKRLGLQESSHREWGQEMVGRMGHAGVYTRNSGSQ"
 polyA_signal 2064..2069
 polyA_site 2090
 BASE_COUNT 462 a 557 c 579 g 506 t
 ORIGIN
 Query Match 26.3%; Score 432.4; DB 9; Length 2104;
 Best Local Similarity 99.8%; Pred. No. 5.9e-102;
 Matches 433; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1211 cccgtgcttctgcatcttctcctcttgacagcctcctgctgtttatctgtgttgaagaa 1270
 |||||||
 Db 1 CCCGTGCTTTGTGCATCTTCTCCTCTGCAGCTCCCTGCTGTTATCTGTGTGGAAAGAA 60
 |||||||
 QY 1271 atgtgctcactctgttactcttcattgcaagagcgtttattcttggagcctttccaagcgg 1330
 |||||||
 Db 61 ATGTGCTCATCTGTGTTACTCTTTCATTGCCAAGAGCGTTATTCTGTGAGGCTTTCACGCG 120
 |||||||
 QY 1331 catatgtttacacacacctgaggtctacccacgcaacgcgccctcgccctggcctggcgcacct 1390
 |||||||
 Db 121 CATATGTTTACACACTGAGTGCTTACCCACGGCAACCGGGCCCTCGGCTGGGCACCT 180
 |||||||

1391	QY	gcagcgcatggcaagatggttgctcctcactccgttccatgccacagtgtgatctgg	1450
Dd			
181	GcAGCGCATGGCAGAGTGGGTCTCTCATCACTCGTTTCATGCCCCAGGTGATCTGG	240	
1451	aatcctctgtgtacctgaactctgcccagtttacagtgtaactgctgctcctggctgcctgg	1510	
Dd			
241	AATCCTCTGTGTACCTGACTCTGCACGTTTTACAGTGCTGCTGCCTCTGCTGCTGCCCTGG	300	
1511	cctcctgcttttggccattgagacaaacgagagacgagagatgcaggagtcaggaccacggg	1570	
Dd			
301	CCTCCTGCTTTTTTCCCATTTGAGACCAGGCGCCAGGACATGACAGAGTCCAGCCACCGG	360	
1571	agtggggcaggagatggttcgcccagaagaatgcacggtgcaggtgttaccaggtctgaact	1630	
Dd			
361	AGTGGGCGCAGGAGATGGTCGCCCGAGGAATGCACGCTGCAGGTGTTACCAAGTTCGAACT	420	
1631	ctggctctcaggaa	1644	
Dd			
421	CTGGCTCTCAGGAA	434	
RESULT	3		
AC087893		212183 bp	HTG
LOCUS	Homo sapiens chromosome 12q clone RP11-423G4, WORKING DRAFT		
DEFINITION	SEQUENCE, 18 unordered pieces.		
ACCESSION	AC087893		
VERSION	AC087893.7 GI:14328965		
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	1 (bases 1 to 212183)		
	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,		
	Bentons,J., Blumage,K., Blankenburg,K., Bonnin,D., Bouckey,J.,		
	Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,		
	Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,		
	Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,		
	Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,		
	Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,		
	Davy-Carroll,L., Dederich,D.A., Delaney,K.K., Delgado,O.,		
	Denn,A.Y., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,		
	Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,		
	Ethal,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,		
	Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,		
	Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,		
	Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,		
	Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,		
	Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,		
	Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,		
	Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,		
	Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,		
	Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,		
	Louisgeed,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,		
	Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,		
	Martinez,E., Massey,E., Mathiney,E., McLeod,M.P., Meador,M.,		
	Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,		
	Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,		
	Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkenwo,S.,		
	Ogun,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,		
	Peery,J., Perez,L., Peters,L., Pickens,K., Primus,E., Pu,L.L.,		
	Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,		
	Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,		
	Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,		
	Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,K., Tameris,K.,		
	Tang,H., Tansey,J., Taylor,C., Taylor,T., Telirod,B., Thomas,N.,		
	Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,		
	Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,		
	Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,		
	Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,		

[illegible]

RESULT	5
AC004642	DNA
LOCUS	INV 01-MAY-1998
DEFINITION	Drosophila melanogaster DNA sequence (P1S DS00543 (D193) and DS02867 (D200)), complete sequence.
ACCESSION	AC004642 AC004302 AC003551 AC003552 AC003553 AC003554 AC003555 AC003556 AC003557 AC003558 AC004308 AC003575 AC003848 AC003849 AC003576 AC003577 AC003578 AC003579 AC003580 AC003581 AC003582 AC003583 AC003584 AC003585 AC003586 AC003587 AC003588 AC003589 AC003590 AC003591 AC003592 AC004642.1 GI:3097827
VERSION	HFG.
KEYWORDS	Drosophila melanogaster (Subclones in Sac from P1 clones DS00543 (D193) and DS02867 (D200)) DNA.
SOURCE	Drosophila melanogaster
ORGANISM	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila
REFERENCE	1 (bases 1 to 148432)
AUTHORS	Celniker,S.E., Aghavani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B., Lomotan,M.A., Mazda,P., Mok,M.S., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,D., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.
TITLE	Sequencing of Drosophila chromosome 2R, region 60B1-60B10 unpublished (1997)
JOURNAL	2 (bases 1 to 148432)
REFERENCE	Celniker,S.E., Aghavani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B., Lomotan,M.A., Mazda,P., Mok,M.S., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,D., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.
TITLE	Direct Submission
JOURNAL	Submitted (01-MAY-1998) Berkeley Drosophila Genome Project, MS 64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, US
COMMENT	Sequence submitted by: Berkeley Drosophila Genome Project Lawrence Berkeley National Laboratory, MS 64-121 Berkeley, CA 94720 For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (http://fruitfly.berkeley.edu/sequence-archive.html) or send email to drosophila@mc.lbl.gov . Library locations: 63_6, 83_30. Location/Qualifiers 1 . 148432 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /chromosome="2R" /map="60B1-60B10" /clone="P1S DS00543 (D193) and DS02867 (D200)" /note="d193 extends from bp 1 to bp 83271 and d200 extends from bp 81802 to bp 148432."
FEATURES	source
BASE COUNT	38713 a 34662 c 35252 g 39805 t
ORIGIN	
Query Match	5.4%; Score 89.4; DB 3; Length 148432;
Best Local Similarity	59.1%; Pred. No. 1.8e-12;
Matches	153; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
Oy	196 gatactttcattggtgaagatgcagtgaagccattggcttggaaaatttcagtggaaag 255
Db	75185 GATACCCTACACTGTTACAGCGCATCAATGCCCTTCGGTTCGGATGTTCACGTCAG 75244
Oy	256 ctgtctgtcttcactggcttgcttgctgatgcattgcattggaggatgatccttcagc 315

RESULT 11
AC067757/c

LOCUS AC067757 62204 bp DNA HTG 27-APR-2000
DEFINITION Homo sapiens chromosome 2 clone RP11-423G3 map 2, LOW-PASS SEQUENCE
SAMPLING.

ACCESSION AC067757
VERSION 1 GI:7651806
KEYWORDS HTGS_PHASE0.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 62204)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 2, clone RP11-423G3
2 (bases 1 to 62204)

REFERENCE

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Glinde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karalas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Melidrim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Olivier,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Snit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE

JOURNAL

COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 19993
Center clone name: 423_G_3

* NOTE: This record contains 78 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 730: contig of 730 bp in length
* 731 830: gap of 100 bp
* 831 1522: contig of 692 bp in length
* 1523 1622: gap of 100 bp
* 1623 2322: contig of 700 bp in length
* 2323 2422: gap of 100 bp
* 2423 3119: contig of 697 bp in length
* 3120 3219: gap of 100 bp
* 3220 3925: contig of 706 bp in length

* 3926 4025: gap of 100 bp
* 4026 4710: contig of 685 bp in length
* 4711 4810: gap of 100 bp
* 4811 5499: contig of 689 bp in length
* 5500 5599: gap of 100 bp
* 5600 6293: contig of 694 bp in length
* 6294 6393: gap of 100 bp
* 6394 7102: contig of 709 bp in length
* 7103 7202: gap of 100 bp
* 7203 7920: contig of 718 bp in length
* 7921 8020: gap of 100 bp
* 8021 8722: contig of 702 bp in length
* 8723 8822: gap of 100 bp
* 8823 9516: contig of 694 bp in length
* 9517 9616: gap of 100 bp
* 9617 10317: contig of 701 bp in length
* 10318 10417: gap of 100 bp
* 10418 11127: contig of 710 bp in length
* 11128 11227: gap of 100 bp
* 11228 11922: contig of 695 bp in length
* 11923 12022: gap of 100 bp
* 12023 12715: contig of 693 bp in length
* 12716 12815: gap of 100 bp
* 12816 13518: contig of 703 bp in length
* 13519 13618: gap of 100 bp
* 13619 14318: contig of 700 bp in length
* 14319 14418: gap of 100 bp
* 14419 15123: contig of 705 bp in length
* 15124 15223: gap of 100 bp
* 15224 15919: contig of 696 bp in length
* 15920 16019: gap of 100 bp
* 16020 16721: contig of 702 bp in length
* 16722 16821: gap of 100 bp
* 16822 17533: contig of 712 bp in length
* 17534 17633: gap of 100 bp
* 17634 18322: contig of 689 bp in length
* 18323 18422: gap of 100 bp
* 18423 19120: contig of 698 bp in length
* 19121 19220: gap of 100 bp
* 19221 19932: contig of 712 bp in length
* 19933 20032: gap of 100 bp
* 20033 20735: contig of 703 bp in length
* 20736 20835: gap of 100 bp
* 20836 21544: contig of 709 bp in length
* 21545 21644: gap of 100 bp
* 21645 22359: contig of 715 bp in length
* 22360 22459: gap of 100 bp
* 22460 23153: contig of 694 bp in length
* 23154 23253: gap of 100 bp
* 23254 23941: contig of 688 bp in length
* 23942 24041: gap of 100 bp
* 24042 24741: contig of 700 bp in length
* 24742 24841: gap of 100 bp
* 24842 25541: contig of 700 bp in length
* 25542 25641: gap of 100 bp
* 25642 26348: contig of 707 bp in length
* 26349 26448: gap of 100 bp
* 26449 27069: contig of 621 bp in length
* 27070 27169: gap of 100 bp
* 27170 27873: contig of 704 bp in length
* 27874 27973: gap of 100 bp
* 27974 28665: contig of 692 bp in length
* 28666 28765: gap of 100 bp
* 28766 29463: contig of 698 bp in length
* 29464 29563: gap of 100 bp
* 29564 30231: contig of 668 bp in length
* 30232 30331: gap of 100 bp
* 30332 31031: contig of 700 bp in length
* 31032 31131: gap of 100 bp
* 31132 31820: contig of 689 bp in length
* 31821 31920: gap of 100 bp
* 31921 32597: contig of 677 bp in length
* 32598 32697: gap of 100 bp

Query Match	5.0%	Score 81.8	DB 2	Length 62204
Best Local Similarity	88.1%	Pred. No. 1.8e-10		
Matches	89	Conservative	0	Mismatches 12; Indels 0; Gaps 0;
948	aacttctgctggtggttatatggtttcccaatgcattctcttactacgggtagttct	1007		
1007				
1008				
1009				
1010				
1011				
1012				
1013				
1014				
1015				
1016				
1017				
1018				
1019				
1020				
1021				
1022				
1023				
1024				
1025				
1026				
1027				
1028				
1029				
1030				
1031				
1032				
1033				
1034				
1035				
1036				
1037				
1038				
1039				
1040				
1041				
1042				
1043				
1044				
1045				
1046				
1047				
1048				
1049				
1050				
1051				
1052				
1053				
1054				
1055				
1056				
1057				
1058				
1059				
1060				
1061				
1062				
1063				
1064				
1065				
1066				
1067				
1068				
1069				
1070				
1071				
1072				
1073				
1074				
1075				
1076				
1077				
1078				
1079				
1080				
1081				
1082				
1083				
1084				
1085				
1086				
1087				
1088				
1089				
1090				
1091				
1092				
1093				
1094				
1095				
1096				

1008 actcaccacagaaactcttccaggcaggagatgtctgcggca 1048
|||||
10557 ACTCACCACAGAACTCTTCCAGGCAGGAGATGTCTGCGGCA 48517

RESULT 12
AE008141/c
LOCUS

DEFINITION	Agrobacterium tumefaciens strain C58 circular chromosome, section 199 of 254 of the complete sequence.
ACCESSION	AE008141
VERSION	AE008141.1
KEYWORDS	GI:15157408
SOURCE	
ORGANISM	Agrobacterium tumefaciens. Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae; Rhizobium.
REFERENCE	1 (bases 1 to 11286) Hinkle,G., Slater,S.C. and Goodner,B. Complete Genome Sequence of Agrobacterium tumefaciens C58 (Rhizobium radiobacter C58), the Causative Agent of Crown Gall Disease in Plants
AUTHORS	Unpublished
JOURNAL	2 (bases 1 to 11286) Hinkle,G., Slater,S.C. and Goodner,B.
AUTHORS	Direct Submission
TITLE	Submitted (14-AUG-2001) Cereon Genomics, 45 Sidney Street, Cambridge, MA 02139, USA
JOURNAL	Location/Qualifiers
FEATURES	1. .11286 /organism="Agrobacterium tumefaciens" /strain="C58" /db_xref="taxon:358" 118..1128 /gene="AGR_C_4098" 118..1128 /note="AGR_C_4098" /note="(AP001509) BH0731 unknown conserved protein" /codon_start=1 /transl_table=11 /product="AGR_C_4098p" /protein_id="AAK87995.1" /db_xref="GI:15157409." /translation="MCVMTGETDEKAIVASAEKALSOLAIIRTEVSKVIFGOEKVV QDTLALISGSHALLVGVPGPARKTKLVTLVGLDANRLOFTPLMPSDILGSEVM DQDEIGRRSFRITKGIIFQGLMADEINRASPTOSLQAQMOEYHITMAGOTVLPK PPIVLAQTPLQEQTPIPEALDFLLQVDVDPPELAAREKLTLDITITGASGEAR VIGAEKMETQALITROMPYSDRVDAILISVRSARGHGNKLTDRKNVWVGPPRAGQS LMLTARALYEGRLAPSLDDVHALAEPLVEHRLMALTTFAARAEGMSVRDVTAAALVEQA KN"
gene	1159..2079 /gene="AGR_C_4100"
CDS	1159..2079 /gene="AGR_C_4100" /note="hypothetical protein" /codon_start=1 /transl_table=11 /product="AGR_C_4100p" /protein_id="AAK87996.1" /db_xref="GI:15157410." /translation="MPVASIGQIVDKTPGSEVLARARORAAALVPDCMVEAKRITANTVT AGWRRKRKIGENWOFPRYAGESLSRIDRRSARDODHTYVRDREWEAAHT IWLWA DMSPSMFKSTGLGSVKSEKALVIMLALAEILLDANSGERIGCGIMEPVSAARNAARLA AAIMHAPLTSCMPDSTIMIRGASDIVLIGDFLDKANKMERISFLARRGLRGHVEIAD PAETFFYSGRTETFSPTGEKLVSGRAETITREDYTRAYLARAREALSSSLRRLLGNWIF FHRTDHLASEALVAVHMYLSCAPAQGGHHR"
gene	2076..4886 /gene="AGR_C_4102"
CDS	2076..4886 /gene="AGR_C_4102" /note="hypothetical protein" /codon_start=1 /transl_table=11 /product="AGR_C_4102p" /protein_id="AAK87997.1" /db_xref="GI:15157411." /translation="MSALPFFVTSPIYILGGLLPAIWLRLRTPPRPKAEVPPPLKI LATVKKRETFPSKSPWMTLLRLALAAAVIPALADPVVNPVNNIS IAGSGPLVIVDONS WASAPDERRVATAQALIGDAERADRVSIPTADAEDHAVSGTTTAVLDKLAAPK PLVPDRARTDAITAEALNAPGTLYIADGVQTQADSEAMRTLASLSLSPAEEFIRVGR

CC The orientation of the sequence is from SP6 to T7 end of the P1
CC clone.

CC Genes were identified by a combination of three methods: Gene
CC prediction programs including GRAIL (available by anonymous ftp
CC from arthur.epm.ornl.gov), GeneFinder (Phil Green, University of
CC Washington), Genscan (Chris Burge,
CC <http://gnomic.stanford.edu/~chris/GENSCANW.html>), and NetPlantGene
CC (<http://www.cbs.dtu.dk/netplantgene/cbsnetpgene.html>), searches of the
CC complete sequence against a peptide database and the Arabidopsis
CC EST database at TIGR (<http://www.tigr.org/tldb/at.html>).
CC Annotated genes are named to indicate the level of evidence for
CC their annotation. Genes with similarity to other proteins are named
CC after the database hits. Genes without significant peptide
CC similarity but with EST similarity are named as 'unknown' proteins.
CC Genes without protein or EST similarity, that are predicted by more
CC than two gene prediction programs over most of their length are
CC annotated as 'hypothetical' proteins. Genes encoding tRNAs are
CC predicted by tRNAscan-SE (Sean Eddy,
CC <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are
CC identified by RepeatMasker (Arian Smit,
CC <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

XX Key Location/Qualifiers

FT source 1. 81020
FT /chromosome="III"
FT /db_xref="taxon:3702"
FT /organism="Arabidopsis thaliana"
FT /cultivar="Columbia"
FT /clone="MGH6"
FT join(<426..518,645..>1034)
FT /gene="MGH6.1"
FT join(426..518,645..1034)
FT /codon_start=1
FT /db_xref="SPTREMBL:O9M7W6"
FT /note="unknown protein"
FT /gene="MGH6.1"
FT /protein_id="AAF35940.1"
FT /translation="MMHERPLPANTDIPGGPHSNPTPTSLHAEVWTLCRIFKRTVS
FT SKRYTDWRELKANGKRDQSNYQAYINFGDNESSTNVNVRGKGYERSVQFL
FT QPTPQHQNPQLMDTHDVSFQHFSDNIHETWPELRSVVEAFPPSFLS"
FT complement(2496..2802)
FT /note="ATTIR16T3A/ATTIR16T3A ATTIR16T3A, a non-autonomous
FT DNA transposon - a consensus."
FT complement(3046..3122)
FT /rpt_family="(GAA)n"
FT join(complement(13663..>13709),complement(13169..13565),
FT complement(<12513..13076))
FT /gene="MGH6.3"
FT join(complement(13663..13709),complement(13169..13565),
FT complement(12513..13076))
FT /codon_start=1
FT /db_xref="SPTREMBL:O9LDD1"
FT /note="hypothetical protein"
FT /gene="MGH6.3"
FT /protein_id="AAF35941.1"
FT /translation="MAVEAHNLNLPSSNREMIHPVEASGVVNTOMRYCTVPTNPV
FT EQQSLFNPYINISPDRLVHOSMKPTIQSVSSSLTFNSDNVDFRVPSSKRREES
FT VYINISATQIQNPTDPLMFLGQDLSSNVQQHFDIDRLISNHVERMEIEERKKTQ
FT GRRIVEAQ6QMLTKRAKDDEINLHIGKLEKVKSLCVENQIWRDVAOSNEATVN
FT ALRSLQQVLAVERNRWEEPTVADDAQSCGSDGDEEERMKLAGAODTKMCR
FT VGMSCSCGKEASVLLPCRMCLCSVCGSSLNTPICKSPKTAASHVNLSS"
FT complement(13950..14024)
FT /rpt_family="(TAAA)n"
FT join(<18851..19090,19255..19381,19476..19614,19697..19785,
FT 19881..>20002)
FT /gene="MGH6.4"
FT join(18851..19090,19255..19381,19476..19614,19697..19785,
FT 19881..20002)
FT /codon_start=1
FT /db_xref="SPTREMBL:O9LDY9"
FT /note="hypothetical protein"
FT /gene="MGH6.4"

FT /protein_id="AAF35942.1"
FT /translation="MASTGLTVAGALLAGDFRLPAVSSSLIPKRTSSSLSCSLNRDLSS
FT PYNCWRLSRGKILTSLSNKRKFAGVKEADGFLSNVEDTDEMFDLFLNKGKRVFKS
FT TDVKSPTAEYDDDAEALAFALAKVASDVKAGDIKVLVFKPLVWTFFIATAFSRP
FT QIDAIGSRMRDLAEKYGKVGANGVKPNSWTLDDFGDVVIHLFLPPORTFYNLDFYGN
FT AMOIELPFEDQSOPRN"
FT join(complement(21254..22367),complement(<20664..21091))
FT /gene="MGH6.5"
FT join(complement(21254..22298),complement(20664..21091))
FT /codon_start=1
FT /db_xref="SPTREMBL:O9M7W5"
FT /note="unknown protein"
FT /gene="MGH6.5"
FT /protein_id="AAF35943.1"
FT /translation="MMAGNOMPTLGRVVKVCDLPVSEGLPSDSYKLAVTTLSSLSAQVSA
FT AIIQFPASDGLALLRSGLDSARIYFQHRDYSYPATNMIHTNDISQWCKTSGYVADPQSWQ
FT ESEYERPLGTTPESNSMEFPAGLPIFALIGKAARVVIDAIGFYLNLRSCPTTEILID
FT NVPLRNCESVSSVSLVCCYARPSFGAQRHSLTEDEOLILYSDHDLKSLISFVKSD
FT KAGLHTRMGQWILVDYDLGPQEAHVYVPGALYQATAGYVSPAVHRTDLNSLQSGISG
FT RFLAFKLMKPSNTLSCSEMPRAAGVGAQFQLPVSDVDFMQRSHSNDLFFNROTLOS
FT FIVQSDGSMKQLKRRKSDSRCKPLPPLSRKRLLEAQRVLKERVQETADKKGILKRF
FT NLKCECNHNMVNSPCANIRREIGWPHGVFPVPHDLPNKAIGFLETPEGWSETHDM
FT EFSLSETAOCNQHVNTCNHSSPNNTPPFAET"
FT join(complement(24815..>25204),complement(24582..24731),
FT complement(24256..24496),complement(23677..24177),
FT complement(<23184..23578))
FT /gene="MGH6.6"
FT join(complement(24815..25204),complement(24582..24731),
FT complement(24256..24496),complement(23677..24177),
FT complement(23184..23578))
FT /codon_start=1
FT /db_xref="SPTREMBL:O9LDF8"
FT /note="unknown protein"
FT /gene="MGH6.6"
FT /protein_id="AAF35944.1"
FT /translation="MPOYSGTGOHCFTAAFSYFWSPTSSRLSNAEERANVFSNLOKEE
FT DDDDEYSPVSTPEKGORATLLELMTIRAFHSKMLRCYSLTGTCGFRIRRGVLTIDIP
FT AIIVFSRKVHKWLSPLQCLTALEGAGGIVWCVDVDFVSEYFGEPOHQPTKOTFTTD
FT IVDHLQSDPFIGSSQVSOETCGTIGAIIVCSQDQVGVFVNRHVAVNLDYPSOKM
FT FHLPLPAGVYLGAVERATSFITDMLFGIFAGTNPETFEVRADGAFIPPADYDILSR
FT VTTSGVGVEIGEIVKAEIQSPVSGVGVKGVKSSGLTTGTGLAYALEYNDERGV
FT CFTLDFLVCGENHRSFPLEGDSGLIVMKGEKARPIGITWGCTGSRGLKLVGECGP
FT ESWTGVDLGRLLTHLQDLITTDDEGLKAAVEORAAITTCMSMWADSSPPYVNLKKE
FT KRSPERLEASLGPLQVQHIDLEERLETGKGAPVSEHQPMPTFSGQCASASWPTARED
FT LVAGFTNGSCDGLCVGLKIGDQKRRRTQVTRNERMPAE"
FT complement(26470..26521)
FT /rpt_family="(GA)n"
FT />26932..>27192
FT /gene="MGH6.7"
FT 26932..27192
FT /codon_start=1
FT /db_xref="SPTREMBL:O9LE44"
FT /note="hypothetical protein"
FT /gene="MGH6.7"
FT /protein_id="AAF35945.1"
FT /translation="MAKNKDDIKYATAQAALSEDAIRVSKYKHTPLEGGKIAESEPEVE
FT LFSSARIEKKEQSAASGQTOIQRDIDKIKGTRTDDSPR"
FT complement(<31715..>32860)
FT /gene="MGH6.8"
FT complement(31715..32860)
FT /codon_start=1
FT /db_xref="SPTREMBL:O9LD39"
FT /note="unknown protein"
FT /gene="MGH6.8"
FT /protein_id="AAF35946.1"
FT /translation="MASGCVKNVGTSPSHSWTSKMSLTRESQPLAPALNEPDVDDFE
FT FLEDDPVTMLSADELFSDGLVPLKFTVYPEEKPTTSVTVAVKPCRLEMEISGV
FT DPLYSPRAPRCTVRWRLEGLARLAKTQOEASASSRSLSSSPNPKTASFRFLNRS
FT SKSTAQPSHPPPKSDIDLESSTSISSRSLSSSSSGHEDDLPLRLSLDLDNKPG
FT TPNPFARSAHHHHHLLHNQPKRPRKTRQVDESTESSIESRVMTVPSDFRLNLSKI
FT VFHGLRSSSSSPCNFTGGPRMKLHHGMRSHSNVRITPVNLVPSLSRSGPKSGLEFFG

Query Match 4.5%; Score 74.2; DB 24; Length 81020;

Result No.	§			ID	Description
	Query	Match	Length		
1	1644	100.0	2562	21	Human organic cation
2	441.2	26.8	1906	21	Rat organic cation
3	359.8	21.9	401	21	Human secreted exp
4	295.4	18.0	480	21	Human 5' EST iso
5	66.2	4.0	3580	22	Human secreted pro
6	66.2	4.0	4366	22	Human brain cell s
7	63	3.8	2268	22	P. putida oxygenase
8	49	3.0	1356	22	C glutamicum codin
9	49	3.0	30940.0	22	C glutamicum codin
10	46.6	2.8	1662	19	Human liver cell c
11	46.6	2.8	1888	19	Human liver cell c

XX Human; organic cation transporter-like protein; OCT1p; transporter;
XX transmembrane; nootropic; neuroprotective; neuroleptic; anticonvulsant;
KW antiParkinsonian; antidepressant; cellular process; cell proliferation;
KW screen; treatment; prevention; diagnosis; neurodegenerative disorder;
KW Parkinson's; Huntington; ALS; amyotrophic lateral sclerosis;
KW CNS disorder; central nervous system; schizophrenia; depression;
KW behavioural; sleep disorder; eating disorder; Alzheimer's; ss.
XX
OS Homo sapiens.

XX	Key	Location/Qualifiers
XX	CDS	185..1831
FF		/*tag= a
FT		/product= "oculp"
FT		
XX		
XX	WO200000633-A1.	
XX		
XX	06-JAN-2000.	
PD		
XX		
XX	29-JUN-1999;	98WO-US14880.
XX		
XX	30-JUN-1998;	98US-O107932.
XX		
XX	(MULTI-)	MILLENNIUM PHARM INC.

XX Goodearl AJ, Glucksmann MA;
PI WPI; 2000-137069/12.
DR P-PSDB; RAY44633.

XX New nucleic acid encoding human organic cation transporter-like
PT protein, used for prevention, treatment and diagnosis of e.g.
PT neurological, behavioural or sleep disorders -
XX

XX Claim 2; Fig 1; 100pp; English.

The present sequence is a cDNA encoding human OCT1p (organic cation transporter-like protein), a member of the superfamily of sugar and other transporter molecules that have 12 transmembrane domains. The sequence is derived from a human foetal brain cDNA library. The protein is highly expressed in brain tissue and has nootropic, neuroprotective, neurolitic, anticonvulsant, antiParkinsonian, antidepressant activities. The present sequence is used to regulate a variety of cellular processes e.g. cell proliferation, differentiation and survival, screen OCT1p modulators and detect mutation in OCT1p gene. OCT1p modulators can be used to treat or prevent chronic neurodegenerative disorders (e.g. Alzheimer's, Parkinson's, Huntington's and ALS), CNS disorders (e.g. schizophrenia, panic, depression), behavioural, sleep and eating disorders.

XX Sequence 2562 BP; 520 A; 682 C; 749 G; 609 T; 2 other;

SQ

Query Match 100.0%; Score 1644; DB 21; Length 2562;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 1644; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Qy 1 atgagaggacttattccagtcaaggcagctgcggatgtgaattcccgacacagc 60
|||||
Db 185 atgagaggacttattccagtcaaggcagctgcggatgtgaattcccgacacagc 244
|||||

Qy 61 gaagtcgaagtcagagagcacgcgttcagagagcatgaagcccagattgaagg 120
|||||
Db 245 gagatgcgaaggtcagagagcacgcgttcagagagcatgaagcccagattgaagg 304
|||||

Qy 121 gtcaccgtggacctagaggcttgagatgatggggcagctgtgccaaaggagttt 180
|||||
Db 305 gtccacgtjggacctagaggcttgagatgatggggcagctgtgccaaaggagttt 364
|||||

Qy 181 gccaatccccacqatgatacttcattggtggagaatgcagtcggagccatttgttgg 240
|||||
Db 365 gccaatccccacqatgatacttcattggtggagaatgcagtcggagccatttgttgg 424
|||||

Qy 241 aaatttcagtgaagctgctcttcactggttagcttgatgagctgagtcgcatggag 300
|||||
Db 425 aaatttcagtggagctgctcttcactggttagcttgatgagctgagtcgcatggag 484
|||||

Qy 301 atgatctctcagcactcgtgccacaagctgcattcgagtggaaggtctccaagctgg 360
|||||
Db 485 atgatctctcagcactcgtgccacaagctgcattcgagtggaaggtctccaagctgg 544
|||||

Qy 361 cagtggtcattctgacctcgtgcttttatgcatgatgtccagctccacgctctgg 420
|||||
Db 545 cagtggtcattctgacctcgtggttttatgcatgatgtccagctccacgctctgg 604
|||||

Qy 421 ggaaatatctcagaccagtcacggcagaaaacagggtctaagatcagcgtgtctgagact 480
|||||
Db 605 ggaaatatctcagaccagtcacggcagaaaacagggtctaagatcagcgtgtctgagact 664
|||||

Qy 481 ctgtactatggactcttagtcatttgcgcccttgtatagcttgactctgtgtctcgg 540
|||||
Db 665 ctgtactatggactcttagtcatttgcgcccttgtatagcttgactctgtgtctcgg 724
|||||

Qy 541 ggcctgtgtggcttogggatcggagaggttcccccagtcggtgacgctgtatccgagttc 600
|||||
Db 725 ggcctgtgtggcttogggatcggagaggttcccccagtcggtgacgctgtatccgagttc 784
|||||

QY	601	cttcccatgaaagccagagctaaatgtattttgctgattgaagatattcglggccaatcgag	666
DB	785	cttcccatgaaagccagagctaaatgtattttgctgattgaagatattcglggccaatcgag	844
QY	661	acagtgctcgaagctcgtcctggctgtgttcgtgagcccaagcctggcgtggcgcttggctg	720
DB	845	acagtgctcgaagctcgtcctggctgtgttcgtgagcccaagcctggcgtggcgcttggctg	904
QY	721	ctcctcctcagctgtcccgctctcctccttctgcggtgctgtgtttcttgctgctgctgaa	780
DB	905	ctcctcctcagctgtcccgctctcctccttctgcggtgctgtgtttcttgctgctgctgaa	964
QY	781	agtgcgaagtatgatgtctgtcagaggaaccaggaaggaatcgccaccttlaaagag	840
DB	965	agtgcgaagtatgatgtctgtcagaggaaccaggaaggaatcgccaccttlaaagag	1024
QY	841	atagcaactgaaacccagagctccctcctcctgcgcgtggggaactcatctccagacagaa	900
DB	1025	atagcaactgaaacccagagctccctcctcctgcgcgtggggaactcatctccagacagaa	1084
QY	901	gaccgaggaacaaatgagggaacttttccacacccattttagatggacaactttgctgctg	960
DB	1085	gaccgaggaacaaatgagggaacttttccacacccattttagatggacaactttgctgctg	1144
QY	961	tggtttatatggttttcccaatgcatctcttttaacgggttagttctactcaccacagaa	1024
DB	1145	tggtttatatggttttcccaatgcatctcttttaacgggttagttctactcaccacagaa	1208
QY	1021	ctcttcaggcagagagatgctcgcgcactctccagtcggaagaagcgtgtagagggcaaaa	1088
DB	1205	ctcttcaggcagagagatgctcgcgcactctccagtcggaagaagcgtgtagagggcaaaa	1268
QY	1081	tgcagctggcctgcgagtaacctgagtgaaagagattacatgaactgctgtgagaccacc	1144
DB	1265	tgcagctggcctgcgagtaacctgagtgaaagagattacatgaactgctgtgagaccacc	1328
QY	1141	ctctctgagtttcagagtgctctgtgactctgtggattattgacgcgctggggcgcaag	1208
DB	1325	ctctctgagtttcagagtgctctgtgactctgtggattattgacgcgctggggcgcaag	1388
QY	1201	aagaccatggccctgtgctttgtcaatctctctctctgcagcctcctgctgtttatctgt	1268
DB	1385	aagaccatggccctgtgctttgtcaatctctctctctgcagcctcctgctgtttatctgt	1448
QY	1261	gttggaagaaatgtgctcaactctgtactcttcatttcaagagcgtttattcttgaggcc	1328
DB	1445	gttggaagaaatgtgctcaactctgtactcttcatttcaagagcgtttattcttgaggcc	1508
QY	1321	tttcaacggcgcatgttttacacacctgaggtctaccccggaacgagggccctcgcc	1388
DB	1505	tttcaacggcgcatgttttacacacctgaggtctaccccggaacgagggccctcgcc	1568
QY	1381	ctgggcaactgcagcgcatggcgaagtggtgctctcatcactcgttcatgcgccag	1448
DB	1565	ctgggcaactgcagcgcatggcgaagtggtgctctcatcactcgttcatgcgccag	1628
QY	1441	gtgatgtggaactctgtgtactgaactcgtgcagtttacagtggctgctgctcctg	1508
DB	1625	gtgatgtggaactctgtgtactgaactcgtgcagtttacagtggctgctgctcctg	1688
QY	1501	gctgcctggcctcctgctttttgccatttgagaccacaaagcgagagactgcagagatcc	1568
DB	1685	gctgcctggcctcctgctttttgccatttgagaccacaaagcgagagactgcagagatcc	1748
QY	1561	agccacggagatgggcccagagatgctgcggccgaggaatgcacggtgcaggtggttacc	1628
DB	1745	agccacggagatgggcccagagatgctgcggccgaggaatgcacggtgcaggtggttacc	1808
QY	1621	aggtcgaaactctggctctcaggaa	1644
DB	1805	aggtcgaaactctggctctcaggaa	1828

expressed sequence tags (seSTs), useful for treating various disorders such as autoimmune, infectious, and central nervous system disorders -

Claim 1: Page 679; 803pp; English.

AAA43426 to AAA45925 represent specifically claimed secreted expressed sequence tags (seSTs), isolated from human, mouse, chicken and rat tissue sources. The seSTs can have a range of activities depending on the tissues they were isolated from. The activities include: chemotactic; proliferative; immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal; antiviral; antidiabetic; antiasthmatic; vulnular; antitumor; osteopathic; neuroprotective; neurotropic; antiparkinsonian; antipsoriatic; cerebroprotective; anticonvulsant; and antidepressant. The seSTs can be used for gene therapy and in vaccines. The seSTs are useful as probes for the identification and isolation of full-length cDNAs and genomic DNA molecules which correspond to the seSTs. Proteins encoded by the seSTs are useful in assays for determining biological activity and raising antibodies. They may be useful for treatment of autoimmune disorders (multiple sclerosis, insulin dependent diabetes), allergic conditions (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers, osteoporosis, osteoarthritis, central nervous system disorders (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's disease), tumours, bacterial fungal or viral infections, depression and psoriasis. AAA45926 to AAA45931 represent linker variants which are given in the exemplification of the present invention.

Sequence 401 Bp; 69 A; 119 C; 104 G; 109 T; 0 other:

metabolic disease; developmental disease; cytostatic; immunomodulatory; neuroprotective; gene therapy; cancer; immunopathology; neuropathology.

Homo sapiens.

WO200132927-A2.

10-MAY-2001.

02-NOV-2000; 2000WO-US30396.

04-NOV-1999; 99US-0163508.

(INCYTE) INCYTE GENOMICS INC.

Sornasse T, Seilhamer JJ, Watson GA;

WPI; 2001-291057/30.

New cell and tissue specific polynucleotides useful for diagnosis, prognosis or monitoring of treatments for disorders where the gene is associated with a cancer, immunopathology or neuropathology.

Claim 1: Page 309-310; 327pp; English.

AAH57161 to AAH57576 represent cell and tissue specific polynucleotide sequences (I). (I) can have cytostatic, immunomodulatory and neuroprotective activities, and can be used in gene therapy. (I) and proteins (II) encoded by them are used in high throughput screening assays to select DNA molecules, RNA molecules, peptide nucleic acids, mimetics, peptides, proteins, agonists, antagonists, antibodies or their fragments, immunoglobulins, inhibitors, drug compounds and pharmaceutical agents. Expression of (I) in a sample indicates the differentiation of embryonic stem cells into a tissue selected from brain, heart, kidney, liver, lung, skeletal muscle or pancreatic tissues. (I) and (II) are used to produce an expression profile that defines a metabolic or developmental process, treatment, condition, disease or disorder. The gene profile can be used for diagnosis, prognosis or monitoring of treatments and for investigating a predisposition to a disorder where the gene is associated with a cancer, immunopathology or neuropathology.

Sequence 4366 BP; 916 A; 1187 C; 1195 G; 1068 T; 0 other;

Query Match 4.0%; Score 66.2; DB 22; Length 4366; Best Local Similarity 47.3%; Pred. NO. 1.8e-08; Matches 234; Conservative 0; Mismatches 258; Indels 3; Gaps 1;

QY 231 tggcttgaaatttcagtgaaagctgtctgtctcactgtctgttgatgtgtga 290
DB 911 tggccacggccgtccagtgacacgtgattttgtgttgctggcgtgagcgtcga 970
QY 291 tgcattggagatgatctcagatccatcggacacacagctgcattgcagtgaggct 350
DB 971 cggctggaggtcttgggtggcttgcgtgccagcgtcgagaaacatgtgct 1030
QY 351 cccaaagtgcagtgatcactgacccgtgtctttagcatgatccagctc 410
DB 1031 gtccagctccaaacaaaggatcagtgactatctatcctcgttgatgtggagc 1090
QY 411 cagctctggggaatatctcagaccagtcagcggcaggaaaacagggcgaagatcagct 470
DB 1091 ctctctggggaggtcgtgacccgtggtggtcggagcagtgctgtcctatcgtc 1150
QY 471 gctggagacttactatggatccttagtcatttgcgccgtgtagctgactct 530
DB 1151 ctcaagtcacagcgtcttgcctctctctcattttgtccagggttaaggcaattct 1210
QY 531 ggtgctccggcgtggtggttcggatccggagg ---agttcccccagctggtgacgct 587
DB 1211 ctctgcgcactcttctcgtgggttggtgaggggtccatcccccattgtcttctcta 1270

QY 588 gtatcccgaggtctctccatgaaagccagagcagataatgtatttctgtgattgaggtatt 647
DB 1271 ttctccgagttcttgcccgaggaagaaagcagggagcatttgagctgctgcatgtt 1330
QY 648 ctgggcccacatcgagagtgctcagagtgctcctcgtgctgtgttcgtgagcggcagcctggg 707
DB 1331 ttggatgattggcgtgtacgcagctgctatgagcctgggcccacatcccccacatagg 1390
QY 708 ctggcgttggtgct 722
DB 1391 gtggagtttccagat 1405

RESULT 7

AAF26353 standard; DNA; 2268 BP.

XX AAF26353;

XX AAF26353;

XX 02-MAY-2001 (first entry)

DE P. putida oxygenase encoding DNA ORF04677.

XX Oxygenase; transgenic plant; detection; amino acid production;
KW vitamin production; steroid production; dialcohol production;
KW oxidized aromatic compound production; dialdehyde production;
KW optically active epoxide production; dicarboxylic acid production;
KW bioconversion; ORF04677; ds.
XX Pseudomonas putida.

XX WO200107629-A2.

XX 01-FEB-2001.

XX 27-JUL-2000; 2000WO-BP07244.

XX 27-JUL-1999; 99DE-1035087.

XX (TIGR-) TIGR INST GENOMIC RES.

PA (QUIA-) QUIAGEN GMBH.

PA (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.

PA (DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.

PA (MEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.

PI Fraser CM, Venter C, Tuemmler B, Holtheisel J, Dueterhoeft A;

PI Hilbert H, Timmis KN, Moore E, Straetz M, Heim S;

XX WPI; 2001-168556/17.

XX New DNA encoding a bacterial oxygenase, useful e.g. for production of amino acids or vitamins, also derived transgenic plants -
XX Claim 3a; Page 35-36; 47pp; German.

XX This invention describes novel DNA sequences (I) that express products having the biological function of oxygenases. The invention also describes (1) recombinant expression vectors containing (I); (2) prokaryotic and eukaryotic cells transformed or transfected with (I) or the vector of (1); (3) production of oxygenases by culturing cells of (2); (4) (partial) expression products (II) of (1), and synthetic proteins or peptides with the same sequences; (5) mono- or poly-clonal antibodies (Ab) specific for (I); (6) hybridoma cells that produce monoclonal Ab; and (7) transgenic plants that contain cells of (2), (1), and their fragments, are useful for the following: (i) expression of recombinant oxygenases; (ii) useful as probes and primers for detection, isolation and amplification of full-length cDNA sequences; and (iii) used to produce transgenic plants. Expression products are useful for production of amino acids, vitamins, steroids, oxidized aromatic compounds, optically active epoxides, dialcohols, dialdehydes and dicarboxylic acids. The transformed cells that express them are used for bioconversion of aromatic and aliphatic compounds. The oxygenases have very weak substrate specificity so can be used for highly stereo- or

CC regio-specific oxidations of a wide range of substrates, reactions that
are difficult or impossible to do with conventional oxidizing agents.
XX
XX Sequence 2268 BP; 371 A; 713 C; 732 G; 450 T; 2 other;

Query Match 3.8%; Score 63; DB 22; Length 2268;
Best Local Similarity 45.9%; Pred. No. 1,1e-07;
Matches 253; Conservative 0; Mismatches 295; Indels 3; Gaps 1;

Qy 233 gcttggaaaattccagtggaagctgtctgttctcactggcttggttgatgagctgagtg 292
Dy 41 gctttaccctgttccactgagtgatgctgctgctgctgctgctgctgctgctgctg 100
Qy 293 ccatggagatgagctcagcagctcctggcaccacacagctgcatcgagtgagctcc 352
Dy 101 gctatgacctgttcactcagctggtgactgcgctcactcagaaagagtgaggccctga 160
Qy 353 caagctggcaggtggcattgtgacctcgtggtgtctttaggcattgatgtccagctcca 412
Dy 161 cccgtgtgcaggtggtgcactgggcagctatgcgtgttcggcattgatgttcggcgccc 220
Qy 413 cgcctctgggaaatatctcagaccagctacggcaggaagaaacaggcctgaagatcagctgc 472
Dy 221 tggcttcggcagctggccagaccgcatcgggcgcaagaaggcattgccattgtttcg 280
Qy 473 tggagctctgactatggcactccttagctcatttgcgcccgtgtatagctggatctgg 532
Dy 281 cctgttctcggggacactatcctaatgctttgcagcaaacccgagcaggttgcca 340
Qy 533 tgcctcggggcctggtggcttcggatcgagagagtt---cccagtcgggtgacgctgt 589
Dy 341 tctaccgttccatcgccgctggtgctgtggcgctgagtcaccaacgctgtggcactga 400
Qy 590 atgcagattcttccactgaagaccagactaaatgtatttgcattgaggtattct 649
Dy 401 tgaacgaatacgcaccacacgctgctgcagcagcgtgggtggcagctatgttcagtggt 460
Qy 650 gggccatcgagacagtgctcagctgctgctggctggtgtgttcgtgatgccacgctggct 709
Dy 461 attcgtggcgagctgctgctgcgaggtgtcggcattcttcattcgtcgcgctgttgct 520
Qy 710 ggcgtgtgctcactcctcagctgtcccgctcctcctctcttcttgcgtgctgttct 769
Dy 521 gggagtcga-gttcttgcgcagcgggtgccactgctgctgttaccgggtattctctact 580
Qy 770 ggcctgcctgaa 780
Dy 581 acctgcctgaa 591

RESULT 8
ID AAH68313
XX AAH68313 standard; DNA; 1356 BP.
XX
XX
XX AC AAH68313;
XX
XX DT 26-SEP-2001 (first entry)
XX
XX DE C glutamicum coding sequence fragment SEQ ID NO: 3348.
XX
XX KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
XX KW organic acid synthesis; ds.
XX
XX OS Corynebacterium glutamicum.
XX
XX PN EP1108790-A2.
XX
XX PD 20-JUN-2001.
XX
XX PF 18-DEC-2000; 2000EP-0127688.
XX
XX 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.
XX 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
XX Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI: 2001-376931/40.
XX P-PSDB; AAC93094.

Novel polynucleotides derived from Coryneform bacteria, for identifying
mutation point of a gene, measuring expression of a gene, analysing
expression profile or pattern of a gene and identifying homologous gene
-
Claim 8; SEQ ID NO: 3348; 246pp + Sequence Listing; English.

The present invention provides a number of nucleotide and protein
sequences from the Coryneform bacterium Corynebacterium glutamicum. These
are useful for identifying the mutation point of a gene derived from a
mutant of coryneform bacterium, measuring expression amount and
analysing the expression profile or expression pattern of a gene derived
from Coryneform bacterium, and identifying a homologue of a gene derived
from Coryneform bacterium. Coryneform bacteria are useful for producing
amino acids, nucleic acids, vitamins, saccharides and organic acids,
particularly L-lysine. The present invention
in the exemplification of the invention.
Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from the
European Patent Office.

Sequence 1356 BP; 230 A; 388 C; 383 G; 355 T; 0 other;

Query Match 3.0%; Score 49; DB 22; Length 1356;
Best Local Similarity 45.5%; Pred. No. 0.00083;
Matches 215; Conservative 0; Mismatches 255; Indels 3; Gaps 1;

Qy 289 gatgccatggagatgatgtatcctcagcactcctggcaccacagctgcattcgagtgagg 348
Dy 184 gacggctacacacattgtgtacggcaccgtccaaatcgccctggctaaaggagtggaac 243
Qy 349 ctcccaagctggcaggtggcattgtgacctggtgcttctttaggcattgatgtccagc 408
Dy 244 ttaagctctcaacgctgggcaccatcgctccaccgcttcttggcattggcgatcgc 303
Qy 409 tccacgctctgggaaataatctcagaccagctacggcaggaagaaacagggctgaagatcagc 468
Dy 304 gctgtgttcattggtcagactgtcagaccgctggcgccgaaagcagcggtgattggatcc 363
Qy 469 gtgctgtggaactgtactatggcattctagtcatttgcgccctgtatagctggtatc 528
Dy 364 gtgctgattctctcttccaccatgctgtgtcatttgcctccaaacccatgggtgttc 423
Qy 529 ctggtgctccggggcctgggtggcttcggatcgaggaggttccccagtcggtga----cg 585
Dy 424 ggcgttccggtttcatcgcagcgttggtctggcggtggtggccttcagttaatgcg 483
Qy 586 ctgtatgccaggttctctcccatgaagcagcagactaaatgtatttgcgtgattgaggtta 645
Dy 484 atgacctctgatttggttccacgtaagaccatgtcagctggggcagcggtgatgctcg 543
Qy 646 ttctgggccatcgaggacagctgttcaggtcgtcgtcgtgctgttcgtgatgccagctcg 705
Dy 544 ggcgttcccatcggtgtctatcgcgctgtgctggcactgtggtgtgttctctctcg 603
Qy 706 ggcgtggcgtgtgctgctcactctcctcagctgtcccgctctctctctctctctctct 758
Dy 604 gaggagtggggctggtcctcactgttctcactgcgttgatccccactggtgtg 656

RESULT 9

AAH68534	
ID	AAH68534 standard; DNA; 309400 BP.
XX	
AC	AAH68534;
XX	
DT	26-SEP-2001 (first entry)
XX	
DE	C glutamicum coding sequence fragment SEQ ID NO: 7069.
XX	
KW	Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX	organic acid synthesis; ds.
XX	
OS	Corynebacterium glutamicum.
XX	
PN	EP1108790-A2.
XX	
PD	20-JUN-2001.
XX	
PF	18-DEC-2000; 2000EP-0127688.
XX	
PR	16-DEC-1999; 99JP-0377484.
XX	
PR	07-APR-2000; 2000JP-0159162.
XX	
PR	03-AUG-2000; 2000JP-0280988.
XX	
PA	(KYOW) KYOWA HAKKO KOGYO KK.
XX	
PI	Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI	Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX	
PI	WPI; 2001-376931/40.
XX	
DR	
XX	
PT	Novel polynucleotides derived from Coryneform bacteria, for identifying
PT	mutation point of a gene, measuring expression of a gene, analysing
PT	expression profile or pattern of a gene and identifying homologous gene
PT	-
XX	
PS	Disclosure; SEQ ID NO: 7069; 246pp + Sequence Listing; English.
XX	
CC	The present invention provides a number of nucleotide and protein
CC	sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC	are useful for identifying the mutation point of a gene derived from a
CC	mutant of coryneform bacterium, measuring expression amount and
CC	analysing the expression profile or expression pattern of a gene derived
CC	from Coryneform bacterium, and identifying a homologue of a gene derived
CC	from coryneform bacterium. Coryneform bacteria are useful for producing
CC	amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC	particularly L-lysine. The present sequence is a nucleic acid described
CC	in the exemplification of the invention.
CC	Note: the sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from the
CC	European Patent Office.
XX	
SQ	Sequence 309400 BP; 70133 A; 86477 C; 83115 G; 69675 T; 0 other;

Query Match	3.0%	Score 49:	DB 22:	Length 309400;
Best Local Similarity	45.5%	Pred. NO. 0.015;		
Matches 215;	Conservative 0;	Mismatches 255;	Indels 3;	Gaps 1;
QY 289	gatgccatggagatgatcttcagcatctcctgagcaccacagctgcattgcgagtggagg	348		
Db 227907	gacggctacgacctcatgtgtacggcaccgtccaatcgccctggctaagagagtggaac	227966		
QY 349	ctcccaagctggcaggctgggcattgtcgaactcgggtgctcttttgacgcattgatccagc	408		
Db 227967	ttaagctctgaacgctgggcaccatcgcgtccaccgcgctcttttggcatggcgatcggc	228026		
QY 409	tcacgcctctggggaatactctcagaccagtaaggcaggaagaaacaggcgtgaagatcagc	468		
Db 228027	gctgtgtctcattggtcgcactgtcagaccgcgctgggcccgaagaagcgcggtgattggatcc	228086		
QY 469	gtcgtgtggactctgtactatggcatcttagtgcatttgcgcgcgtgtatagctgattc	528		

[illegible]

Query Match 2.8% Score 46.6; DB 19; Length 1662;
Best Local Similarity 47.8%; Pred. No. 0.0045;
Matches 232 Conservative 244; Indels 9; Gaps 3
QV 1083 cagccctgcacgaattaccctgaaagaagattcacatgacttgcctgaaccacct 1142

```

Db 1092 caccctgcacatggcgccaccagcggaacccctcaccctggatttcttactccgctct 1151
QY 1143 ctctgagttccaggtgtcctgtgactctgtgattattgaccgctggtggcgcaagaa 1202
Db 1152 ggtcgaaatcccgggccctcagccctcaccctcaccattgaccgctggtggcgcatcta 1211
QY 1203 gaccatggccctgtgttgcctctctctctctctctctctctctctctctctctctct 1259
Db 1212 cccatggccgtgtcaaatlgttggcgggcgagcctgctgctgctgctgctgctgctgct 1271
QY 1260 tgttgaagaagatgtcctcactctgttactctctca---tttgaagagcgtttattcttg 1316
Db 1272 acctgacctgcactgtttaaacatcataatcatgtgtgttggcggaatgggaatcaccat 1331
QY 1317 aggtcttcaagcgcatatgtttacacacctgaggtctacccacacggaacgcgccct 1376
Db 1332 tgaatacaaatgatctgctggtggaatgctgagctgagctgacccacacatttgcaggaaacct 1391
QY 1377 cggcctgggcacctgcagcgccatggcaagagtggtgtctctcactcactcctgttctatcgc 1436
Db 1392 cggagtgatggtgtgtctcctctgtgacatagtggtggataatcacccctctcatagt 1451
QY 1437 ccagggtgactgaatcctctgtgtactgactctggcaggtttacagtggtgctgctct 1496
Db 1452 ctccaggtgagggaggtgtgcaagccttgcctcctcattttgttccggtgttggcct 1511
QY 1497 cctggctgctgctgctcctgctgttttgcctattgagaccacgagcgaggaactgcagga 1556
Db 1512 gcttgcgcggagtgacgtactctctctca---gagaccacaggggtgcgttgcaga 1568
QY 1557 gtcca 1561
Db 1569 gacca 1573

```

RESULT 11

AAV49558
ID AAV49558 standard; cDNA to mRNA; 1888 BP.

XX AC AAV49558;

XX DT 21-OCT-1998 (first entry)

XX DE Human liver cell clone HP01293 cDNA #1.

XX Transmembrane domain; human; nutrition; cytokine; cell proliferation;
KW differentiation; immune system; stimulator; suppressor; regulator;
KW hematopoiesis; activin; inhibitor; chemolactic; chemokinetic; receptor;
KW haemostatic; thrombolytic; ligand; anti-inflammatory; tumour; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX CDS 90..1745

XX FT /tag= a

XX FT /product= "transmembrane domain containing protein"

XX W09821328-A2.

XX PN 22-MAY-1998.

XX PF 07-NOV-1997; 97WO-JP04056.

XX PR 13-NOV-1996; 96JP-0301429.

XX (PROT-) PROTEGENE INC.

XX PA (SAGA) SAGAMI CHEM RES CENTRE.

XX PI Kato S, Kobayashi M, Sekine S, Yamaguchi T;

XX WPI; 1998-297932/26.

XX DR P-PSDB; AAW64538.

XX Human protein having transmembrane domain - useful for, e.g.
PT research and nutrition
XX Claim 4; Page 141-143; 205pp; English.
XX AAV49550-V49599 are cDNA sequences which encode human proteins containing
CC a transmembrane domain. These proteins can be used for, e.g. research
CC and nutrition, and may have cytokine and cell
CC proliferation/differentiation, immune stimulating/suppressing,
CC hematopoiesis regulating, tissue growth, activin/inhibin,
CC chemotactic/chemokinetic, haemostatic and thrombolytic, receptor/ligand,
CC anti-inflammatory or tumour inhibition activity.
XX SQ Sequence 1888 BP; 365 A; 545 C; 520 G; 458 T; 0 other;

Query Match 2.8%; Score 46.6; DB 19; Length 1888;

Best Local Similarity 47.8%; Pred. No. 0.0048;

Matches 232; Conservative 0; Mismatches 244; Indels 9; Gaps 3;

QY 1083 cagcctggcctgcgagctacctgagtgaggagattacatgactgtgtgaccacccct 1142

Db 1181 catctgcacatggcgccaccagcggaacctctacctgatttcttactccgctct 1240

QY 1143 ctctgagtttccaggtgtcctgtgactctggtgattattgaccgctggcgcaagaa 1202

Db 1241 ggtcgaaatccggggccttcacagccctcaccattgaccgctggcgcaatcta 1300

QY 1203 gaccatggcctgctgttctgtcactctctctctctctctctctctctctctctctctctct 1259

Db 1301 cccatggcctgtcaaatgttggcgggcgagcctgctgctgatttctctc 1360

QY 1260 tgttgaagaagatgtcctcactctgttactcttca---ttgcaagagcgtttattcttg 1316

Db 1361 acctgacvtgcactggttaaacatcataatcatgtgttggcggaatgggaatcaccat 1420

QY 1317 aggtcttcaagcgcatatgtttacacacctgaggtctacccacacggaacgcgccct 1376

Db 1421 tgaatacaaatgatctgctgctggtgaaatgctgagctgtacccacattctcgaggaacct 1480

QY 1377 cggcctgggcacctgcagcgccatggcaagagtggtgtctctcactcactcctgttcacgc 1436

Db 1481 cggagtgatggtgtgtctcctcctctgtgacataggtgggataatcaccctctcatagt 1540

QY 1437 ccagggtgactgtaacccctctgtgtacctgactctggcagtttacagtgctgctgct 1455

Db 1541 ctccaggtgagggaggtctggaagccttgccctcatttctgttgcggtgttgggacct 1601

QY 1497 cctggctgcccctggcctcctctgcttttggccattgagaccacgagcgaggaactgcagga 1556

Db 1601 gcttgcgcgggagtgacgtactctcttcca---gagaccacaggggtcgtcttgcaga 1657

QY 1557 gtcca 1561

Db 1658 gacca 1662

RESULT 12

AAC93874

ID AAC93874 standard; cDNA; 549 BP.

XX AC AAC93874;

XX DT 19-FEB-2001 (first entry)

XX Cat flea hindgut and Malpighian tubule (HMT) cDNA, SEQ ID NO:369.

XX Cat flea; hindgut and Malpighian tubule nucleic acid; HMT;

XX KW flea infestation; vaccine; antiparasitic; therapeutic target;

XX KW diagnosis; detection; ss.

XX OS Ctenocephalides felis.

```
XX PN WO200061621-A2.
XX PD 19-OCT-2000.
XX PF 07-APR-2000; 2000WO-US09437.
XX PR 09-APR-1999; 99US-0128704.
XX PA (HESK-) HESKA CORP.
XX PI Brandt KS, Gaines PJ, Stinchcomb DT, Wisniewski N;
XX DR WPI; 2000-656323/63.
XX PT Flea Malpighian tubule and head and nerve cord tissue derived nucleic
XX PT acids useful for the prevention, diagnosis and treatment of flea
XX PT infestations -
XX PS Claim 26; Page 368; 964pp; English.
XX CC The invention relates to novel cat flea (Ctenocephalides felis) nucleic
XX CC acids which are expressed in hindgut and Malpighian tubule (HMT) tissue
XX CC or head and nerve cord (HNC) tissue. The invention also relates to the
XX CC encoded proteins. The invention additionally encompasses expression
XX CC constructs, recombinant viruses and recombinant cells comprising the
XX CC nucleic acids of the invention, recombinant production of the proteins,
XX CC antibodies against the proteins, a method of identifying inhibitors of
XX CC the proteins, and compositions comprising the inhibitors for
XX CC administration to an animal. The nucleic acids, and the proteins they
XX CC encode may be used in the prevention, treatment and diagnosis of diseases
XX CC associated with flea infestations. For example, the nucleic acids may be
XX CC used to produce an HMT or HNC protein according to standard recombinant
XX CC DNA methodology by inserting the nucleic acids into a host cell and
XX CC culturing the cell to express the protein. The HMT and HNC nucleic acids
XX CC may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect
XX CC and quantitate the presence of cat flea or other homologous nucleic acid
XX CC sequences in samples. They may also be used to study the expression and
XX CC function of the proteins and their role in metabolism. The HMT and HNC
XX CC proteins may be used as antigens in the production of specific
XX CC antibodies, and in assays to identify modulators (agonists and
XX CC antagonists) of HMT and/or HNC protein expression and activity. The
XX CC anti-HMT/HNC protein antibodies and antagonists may also be used to
XX CC downregulate protein expression and activity. The antibodies may also be
XX CC used as diagnostic agents for detecting the presence of flea polypeptides
XX CC in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The
XX CC present sequence represents a cat flea HMT cDNA of the invention.
XX SQ Sequence 549 BP; 162 A; 83 C; 122 G; 182 T; 0 other;

Query Match 2.8%; Score 46.4; DB 21; Length 549;
Best Local Similarity 48.8%; Pred. No. 0.0029;
Matches 125; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

Qy 207 ggtggaagatgcagtggaagccattggttggaaaatttcagtggaagctgtctgtctt 266
Db 101 gtttgaagatgcgttggcaactcacaggttttggaaaatttcattaccttcttctgtgctg 160
Qy 267 cactgctgttgcgtggtgctgatgccatggagatgatcatcctcagcatcctggcacc 326
Db 161 gagtggatgcattagtattgttttgatggaaactcttggaaatgatttgcgttcc 220
Qy 327 acagctgcattgcagtgagggtccccaagctggcagtggtcattgctgacctcgttgg 386
Db 221 ttacgacaatgtatcttggaaatlaacacaaacaaaggaataatttaacgctatagc 280
Qy 387 cttttaggcattgattccactccactcctctctgggaaataatctcagaccagtcaggcag 446
Db 281 tttttaggtattacaagcagcttccacatttattgggggttttttagccgatacagagagag 340
Qy 447 gaaacagggtcgaag 462
| | | | |
```

```
Db 341 gcggaagtgattatg 356

RESULT 13
AAC95401
ID AAC95401 standard; cDNA; 1590 BP.
XX AC AAC95401;
XX DT 19-FEB-2001 (first entry)
XX DE Cat flea HMT synaptic vesicle 2B-like cDNA ORF, SEQ ID NO:1904.
XX KW Cat flea; hindgut and Malpighian tubule nucleic acid; HMT;
XX KW flea infestation; vaccine; antiparasitic; therapeutic target;
XX KW diagnosis; detection; ss.
XX OS Ctenocephalides felis.
XX PN WO200061621-A2.
XX PD 19-OCT-2000.
XX PF 07-APR-2000; 2000WO-US09437.
XX PR 09-APR-1999; 99US-0128704.
XX PA (HESK-) HESKA CORP.
XX PI Brandt KS, Gaines PJ, Stinchcomb DT, Wisniewski N;
XX DR WPI; 2000-656323/63.
XX DR P-PSDB; AAB29626.
XX PT Flea Malpighian tubule and head and nerve cord tissue derived nucleic
XX PT acids useful for the prevention, diagnosis and treatment of flea
XX PT infestations -
XX PS Claim 1; Page 926-937; 964pp; English.
XX CC The invention relates to novel cat flea (Ctenocephalides felis) nucleic
XX CC acids which are expressed in hindgut and Malpighian tubule (HMT) tissue
XX CC or head and nerve cord (HNC) tissue. The invention also relates to the
XX CC encoded proteins. The invention additionally encompasses expression
XX CC constructs, recombinant viruses and recombinant cells comprising the
XX CC nucleic acids of the invention, recombinant production of the proteins,
XX CC antibodies against the proteins, a method of identifying inhibitors of
XX CC the proteins, and compositions comprising the inhibitors for
XX CC administration to an animal. The nucleic acids, and the proteins they
XX CC encode may be used in the prevention, treatment and diagnosis of diseases
XX CC associated with flea infestations. For example, the nucleic acids may be
XX CC used to produce an HMT or HNC protein according to standard recombinant
XX CC DNA methodology by inserting the nucleic acids into a host cell and
XX CC culturing the cell to express the protein. The HMT and HNC nucleic acids
XX CC may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect
XX CC and quantitate the presence of cat flea or other homologous nucleic acid
XX CC sequences in samples. They may also be used to study the expression and
XX CC function of the proteins and their role in metabolism. The HMT and HNC
XX CC proteins may be used as antigens in the production of specific
XX CC antibodies, and in assays to identify modulators (agonists and
XX CC antagonists) of HMT and/or HNC protein expression and activity. The
XX CC anti-HMT/HNC protein antibodies and antagonists may also be used to
XX CC downregulate protein expression and activity. The antibodies may also be
XX CC used as diagnostic agents for detecting the presence of flea polypeptides
XX CC in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The
XX CC present sequence represents a cat flea HMT cDNA of the invention.
XX SQ Sequence 1590 BP; 482 A; 252 C; 349 G; 507 T; 0 other;

Query Match 2.8%; Score 46.4; DB 21; Length 1590;
Best Local Similarity 48.8%; Pred. No. 0.0051;
Matches 125; Conservative 0; Mismatches 131; Indels 0; Gaps 0;
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 8, 2002, 02:35:33 ; Search time 137.14 Seconds
(without alignments)
2714.962 Million cell updates/sec

Title: US-09-911-667A-3
Perfect score: 1644
Sequence: 1 atdgaggagactattcca.....cgaactctggtcaggaa 1644

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A-COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B-COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A-COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B-COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS-COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53.2	3.2	1885	US-08-501-572-5	Sequence 5, Appli
2	53.2	3.2	1885	US-09-040-444-5	Sequence 5, Appli
3	45.2	2.7	7218	US-08-232-463-14	Sequence 14, Appli
4	45	2.7	1896	US-08-501-572-6	Sequence 6, Appli
5	45	2.7	1896	US-09-040-444-6	Sequence 6, Appli
6	38.6	2.3	1882	US-08-501-572-4	Sequence 4, Appli
7	38.6	2.3	1882	US-09-040-444-4	Sequence 4, Appli
8	37.6	2.3	50937	US-09-428-517-1	Sequence 14, Appli
9	36.6	2.2	7218	US-08-232-463-14	Sequence 3, Appli
10	36.4	2.2	1490	US-08-964-127-3	Sequence 1, Appli
11	36.4	2.2	2460	US-08-964-127-1	Sequence 1, Appli
12	35.6	2.2	3489	US-08-728-323A-1	Sequence 20, Appli
13	35.6	2.2	32207	US-08-770-379-20	Sequence 20, Appli
14	35.6	2.2	32207	US-08-757-669A-20	Sequence 2, Appli
15	35.2	2.1	2496	US-08-073-384C-2	Sequence 2, Appli
16	35.2	2.1	2496	US-08-234-359A-2	Sequence 2, Appli
17	35.2	2.1	2496	US-08-483-043-2	Sequence 2, Appli
18	35.2	2.1	2496	US-08-481-238-2	Sequence 2, Appli
19	35.2	2.1	2496	US-08-471-066B-2	Sequence 2, Appli
20	35.2	2.1	2496	US-08-484-956-2	Sequence 2, Appli
21	35.2	2.1	2496	US-08-757-653-2	Sequence 2, Appli
22	35.2	2.1	2496	US-08-599-491-2	Sequence 2, Appli
23	35.2	2.1	2496	US-08-756-386-2	Sequence 2, Appli
24	35.2	2.1	2496	US-08-823-516-2	Sequence 2, Appli
25	35.2	2.1	2496	US-08-682-853A-2	Sequence 2, Appli
26	35.2	2.1	2496	US-08-759-038-2	Sequence 2, Appli
27	35.2	2.1	2496	US-08-758-314-2	Sequence 2, Appli

Sequence 2, Appli
Sequence 1, Appli
Sequence 17, Appli
Sequence 17, Appli
Sequence 46, Appli
Sequence 3, Appli
Sequence 1, Appli
Sequence 12, Appli
Sequence 12, Appli
Sequence 12, Appli
Sequence 12, Appli
Sequence 5, Appli
Sequence 21, Appli
Sequence 21, Appli
Sequence 21, Appli
Sequence 21, Appli
Sequence 1, Appli
Sequence 749, App

28 35 2.1 2082 4 US-09-440-325A-2
29 34.8 2.1 2102 2 US-08-647-397-1
30 34.4 2.1 289 4 US-09-007-005-17
31 34.4 2.1 289 4 US-09-244-796-17
32 34.2 2.1 2484 4 US-09-276-531-46
33 34 2.1 2515 1 US-08-061-465-3
34 34 2.1 6828 1 US-08-061-465-1
35 33.6 2.0 803 2 US-07-928-611-12
36 33.6 2.0 803 2 US-08-487-811A-12
37 33.6 2.0 803 4 US-09-060-694-12
38 33.6 2.0 803 5 PCT-US93-07370-12
39 33.6 2.0 1610 1 US-08-056-051-5
40 33.6 2.0 1610 1 US-07-928-611-21
41 33.6 2.0 1610 2 US-08-487-811A-21
42 33.6 2.0 1610 4 US-09-060-694-21
43 33.6 2.0 1610 5 PCT-US93-07370-21
44 33.4 2.0 4089 1 US-07-908-245-1
45 33 729 4 US-08-998-416-749

ALIGNMENTS

RESULT 1
US-08-501-572-5
; Sequence 5, Application US/08501572
; Patent No. 6063623
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein Which Effects The
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics And/or Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan,Henderson,Farabow,Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/501,572
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Toohey, Kimberlin M
; REGISTRATION NUMBER: 35,391
; REFERENCE/DOCKET NUMBER: 02481.1453-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4400
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1885 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-501-572-5

Query Match 3.2%; Score 53.2; DB 3; Length 1885;
Best Local Similarity 47.3%; Pred. No. 8.3e-06;
Matches 228; Conservative 0; Mismatches 248; Indels 6; Gaps 2;
Oy 1083 cagcctggcctgcgagtagctacgtgaggaggattacatgagtagctgtgtggaccacct 1142

Db 1164 CATCTTCACATGGCGCCACACGAGGAACTCTACCTGGATTTCCTTACTCCGCTCT 1223
QY 1143 ctctgagttccagggtgctcttgactctgtgagttatgacctgctgggagcaagaa 1202
Db 1224 GGTGAAATCCCGGGGCTTCATAGCCCTCATCACCATTGACCGCTGGCGGCATCTA 1283
QY 1203 gaccatggcctgtgctttgtatctctctctctcctcctcctcctcctcctcctcct 1262
Db 1284 CCCCATGGCGCTGCAAAATTTGTTGGGGGGGAGCTGCCTGCTCATTTTATCTCACC 1343
QY 1263 tggagaagaatgtctcactcgttactcttca---ttgcaagagcglttatttcttgagg 1319
Db 1344 TGACCTCCACTGGTTAAACATCATATCATGTGTGTGGCCGAATGGAATCACCATTGC 1403
QY 1320 ctttcaagcgcatatgtttacacacctgaggtctacacacacacacacacacacacac 1379
Db 1404 AATACAAATGATCTGCCTGGTGAATGTGAGCTGTACCCACCATTCGTACAGAACCTCAG 1463
QY 1380 cctgggcaacctgcaagcgcatggcaagagtggtgtctctcactcactcctcctcctcctc 1439
Db 1464 AGTATGCTGTTCCTCCCTGGCAAGCTTGCCTCATTTTGTGGCGGTGTTGGCGCTGCT 1523
QY 1440 ggtgatgctggaatcctctgtgtacctgtgacgtctggtgagtttacagtggtcgtcctcct 1499
Db 1524 CAGGCTGAGCGAGCTGGCAAGCTTGCCTCATTTTGTGGCGGTGTTGGCGCTGCT 1583
QY 1500 ggtcgcctgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1559
Db 1584 TGCCGGGGGAGTGACGCTACTTCTTCC---AGAGACCAAGGGGACGCTTTGCCAGAGAC 1640
QY 1560 ca 1561
Db 1641 CA 1642

RESULT 2
US-09-040-444-5
; Sequence 5, Application US/09040444
; Patent No. 6063766
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grunden, Dirk
; TITLE OF INVENTION: Transport protein Which Effects The
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics And/or Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040.444
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 1885 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-040-444-5

Query Match 3.2%; Score 53.2; DB 3; Length 1885;
Best Local Similarity 47.3%; Pred. No. 8.3e-06;
Matches 228; Conservative 0; Mismatches 248; Indels 6; Gaps 2;

QY 1083 cagcctggcctgagtagtaccatgagtagagagatlacatgacttgctgtgagcacctc 1142
Db 1164 CATCTTCACATGGCGCCACACGAGGAACTCTACCTGGATTTCCTTACTCCGCTCT 1223
QY 1143 ctctgagttccagggtgctcttgactctgtgagttatgacctcctcctcctcctcctcct 1202
Db 1224 GGTGAAATCCCGGGGCTTCATAGCCCTCATCACCATTGACCGCTGGCGGCATCTA 1283
QY 1203 gaccatggcctgtgctttgtatctctctcctcctcctcctcctcctcctcctcctcctc 1262
Db 1284 CCCCATGGCGCTGCAAAATTTGTTGGGGGGGAGCTGCCTGCTCATTTTATCTCACC 1343
QY 1263 tggagaagaatgtctcactcgttactcttca---ttgcaagagcglttatttcttgagg 1319
Db 1344 TGACCTCCACTGGTTAAACATCATATCATGTGTGTGGCCGAATGGAATCACCATTGC 1403
QY 1320 ctttcaagcgcatatgtttacacacctgaggtctacacacacacacacacacacacac 1379
Db 1404 AATACAAATGATCTGCCTGGTGAATGTGAGCTGTACCCACCATTCGTACAGAACCTCAG 1463
QY 1380 cctgggcaacctgcaagcgcatggcaagagtggtgtctctcactcactcctcctcctcctc 1439
Db 1464 AGTATGCTGTTCCTCCCTGGCAAGCTTGCCTCATTTTGTGGCGGTGTTGGCGCTGCT 1523
QY 1440 ggtgatgctggaatcctctgtgtacctgtgacgtctggtgagtttacagtggtcgtcctcct 1499
Db 1524 CAGGCTGAGCGAGCTGGCAAGCTTGCCTCATTTTGTGGCGGTGTTGGCGCTGCT 1583
QY 1500 ggtcgcctgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1559
Db 1584 TGCCGGGGGAGTGACGCTACTTCTTCC---AGAGACCAAGGGGACGCTTTGCCAGAGAC 1640
QY 1560 ca 1561
Db 1641 CA 1642

RESULT 3
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232.463

RESULT 9
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFELINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:

[illegible]

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 2.2%; Score 36.6; DB 1; Length 7218;
Best Local Similarity 4.6%; Pred.No. 1.3;
Matches 15; Conservative 175; Mismatches 139; Indels 0; Gaps 0;

QY 3 ggagaggactatccagctaaagcagctgcgggtgtgaaattccgtgcacagcgca 62
Db 1370 RR 1311
QY 63 ggtgcaagtcagaggagcacagcttcagagagacatgaagtcagattgaagggt 122
Db 1310 RR 1251
QY 123 ccagctgjcctagagctgtgagctgatgagggcagctgtgccaggaggtlqgc 182
Db 1250 RR 1191
QY 183 caatccacagtgatactttcatgtgaaacacaglgaaagccattggccttggaaa 242
Db 1190 RR 1131
QY 243 attcagtgaaagctgtcttctcactggcttggtgagctggtgatgcatggagat 302
Db 1130 RR 1071
QY 303 gatgactcagcatcctgcacacagc 331
Db 1070 RRRATCGCAAGCTCCCTGCAGCTGCAGC 1042

RESULT 10
US-08-964-127-3
; Sequence 3, Application US/08964127
; Patent No. 6277565
; GENERAL INFORMATION:
; APPLICANT: Grandearl, Andrew David John
; TITLE OF INVENTION: NOVEL GENES ENCODING TRANSPORTER-LIKE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street

CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/964,127
FILING DATE: 06-NOV-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Crews, Ph.D., L. Lee
REGISTRATION NUMBER: P-43,567
REFERENCE/DOCKET NUMBER: 07334/038001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1490 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 492...1349
US-08-964-127-3

Query Match 2.2%; Score 36.4; DB 4; Length 1490;
Best Local Similarity 47.1%; Pred. No. 0.64;
Matches 112; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

QY 351 cccaagctgcaggtgagcattgtgcacctcgggtggtctttagtagcatgatgtccagctc 410
Db 776 CTTGGCTGCGCAGGTGATCTCTGGAGCAGATCTCTTCATCTTGGCTTTGCCCTCCGGCTA 835
QY 411 cacctctggggaatctctcagaccagtcagcaggaagaaacagggctgaagatcagcgtl 470
Db 836 CTTGTTCTGGTTTACCCCGCAGACAGATTTGGCCGCGGATTGTGCTGTGACCTT 895
QY 471 gcttggaactctgtactatgagcatccttagtcatttgcgcctgtgtatagctggatcct 530
Db 896 GGGGCTGTGGGCCCCCTGTGTGAGTAGAGGGGCTGTGAGGGCTCTCCACAGGGCTCAT 955
QY 531 ggtgtcgcgggctggtgggcttcgggagtcgaggaggttccccagtcggtgacgctg 588
Db 956 GGCCTCCGATCTCTTTGGGCTTTCTCTGCTTGGCGGTGTTGACCTGGGTGTCTACCTG 1013

RESULT 11
US-08-964-127-1
; Sequence 1, Application US/08964127
; Patent No. 6277565
; GENERAL INFORMATION:
; APPLICANT: Grandearl, Andrew David John
; TITLE OF INVENTION: NOVEL GENES ENCODING TRANSPORTER-LIKE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804

us-09-911-667a-3.rni

Tue Mar 12 15:54:35 2002

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/964.127
FILING DATE: 06-NOV-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Crews, Ph.D., L. Lee
REGISTRATION NUMBER: P-43,567
REFERENCE/DOCKET NUMBER: 07334/038001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2460 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 498...2057
US-08-964-127-1

Query Match 2.2%; Score 36.4; DB 4; Length 2460;
Best Local Similarity 47.1%; Pred. No. 0.84; 126; Indels 0; Gaps 0;
Matches 112; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

QY 351 cccaaagtcgagcgtgacattgctgacctcggtggtcttlttgagcagatgctccagctc 410
DB 702 CTGGGCTGGCAGGTGATCTGTCGAGCAGATCTCTTCATCTTGGGCTTTGGCTCCGGCTA 841
QY 411 cagcgtctgggaaatatctcagaccagtcgagcaggaagagggcgtgaagatcagcgt 470
DB 842 CCGTGTCTCGGTTACCCGAGACAGATTTGGCCGTCGCGGGGATTTGCTGCTGACCTT 901
QY 471 gctgtgactctgactatgcatctcttagtgcatttgcgcccgtgtatagctggatcct 530
DB 902 GGGGCTGGTGGCCCTCTGGAGTAGGAGGGGCTGCTGCAGGCTCTCCACAGGCGTCAT 961
QY 531 ggtgtctccgggacctggtggttcgagtcgagcagagagttcccgagtcggtgacgtc 588
DB 962 GGCCCTCCGATTCTCTTGGGCTTTCTGCTTGGCGGTGTTGACCTGGGTGCTACCTG 1019

RESULT 12
US-08-728-323A-1
Sequence 1, Application US/08728323A
Patent No. 5948676
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: Immediate Early Protein From Kaposi's
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
TITLE OF INVENTION: Encoding Same And Uses Thereof
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.

ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,323A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSK/SKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3489 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3489
US-08-728-323A-1

Query Match 2.2%; Score 35.6; DB 2; Length 3489;
Best Local Similarity 51.9%; Pred. No. 1.7;
Matches 80; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 76 gaggacgacacggcttcaggagagatgaatccagattgaagggtccacgtggccta 135
DB 1111 GAGGATGACGAGGAGTACGACGAGGAGGATGACGAGGAGGATGACGAGGAGGAGGAGG 1170
QY 136 gagcgtgtgagctgagatgagtggtggtggtggtggtggtggtggtggtggtggtggt 195
DB 1171 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1230
QY 196 gatactttcatgtggaagatgcagtcgagtcga 229
DB 1231 AATGAGGAGGAGGAGGAGGATGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1264

RESULT 13
US-08-770-379-20/c
Sequence 20, Application US/08770379
Patent No. 5849564
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

```

; APPLICATION NUMBER: US/08/770,379
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 52342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3207 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-770-379-20

Query Match 2.2%; Score 35.6; DB 2; Length 32207;
Best Local Similarity 51.9%; Pred. No. 5.6;
Matches 80; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

Qy 76 gagacacacgcgttcaggagagacatgaagtcacagattgaagggtccacgtgggccta 135
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20886 GAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGAGGAG 20827

Qy 136 gagcgttgagctggtgatgagtgaggcagctgccccaaaggagtttgcacatccaccgat 195
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20826 GAGGAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 20767

Qy 196 gatactttcgtgtggaagatgcagtggaagcca 229
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20766 AATGAGGACGAGGAGGATGACGAGGAGGAGGACA 20733

RESULT 15
US-08-073-384C-2
; Sequence 2, Application US/08073384C
; Patent No. 5541311
; GENERAL INFORMATION:
; APPLICANT: Dahlberg, James E.
; APPLICANT: Lyamichev, Victor I.
; APPLICANT: Brow, Mary Ann D.
; TITLE OF INVENTION: SYNTHESIS-DEFICIENT THERMOSTABLE DNA
; TITLE OF INVENTION: POLYMERASE
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/073,384C
; FILING DATE: 04-JUN-1993
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/986,330
; FILING DATE: 07-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Peter G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: FORS-00613
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/705-8410
; TELEFAX: 415/397-8338
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2496 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-073-384C-2

```

Query Match 2.1%; Score 35.2; DB 1; Length 2496;
 Best Local Similarity 62.5%; Pred. No. 1.9;
 Matches 55; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 456 gctgaagatcagcgctgtggtactctgtactatggcatccttagtgcatttgccccgt 515
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1740 GCAGAACATCCCGGNGCGACCCCTCTGGGCCAGCCCATCCGCCGAGCCTTCGTGCCGA 1799

QY 516 gtatagctggatcctctgtgtgtcggggc 543
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1800 GGAGGGCTGGTCTGTGGTCTTGGAC 1827

Search completed: March 8, 2002, 02:39:31
 Job time: 10454 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 8, 2002, 00:47:58 ; Search time 3543.72 Seconds
(without alignments)
4985.174 Million cell updates/sec

Title: US-09-911-667A-3
Perfect score: 1644
Sequence: 1 atgagagagactattcca.....cgaactctggtctcaggaa 1644

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST : *
- 1: em_estfun:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estom:*
 - 5: em_estpl:*
 - 6: em_estba:*
 - 7: em_estro:*
 - 8: em_estov:*
 - 9: em_hic:*
 - 10: gb_est1:*
 - 11: gb_est2:*
 - 12: gb_hic:*
 - 13: gb_gss:*
 - 14: em_gss_fun:*
 - 15: em_gss_hum:*
 - 16: em_gss_inv:*
 - 17: em_gss_pln:*
 - 18: em_gss_pro:*
 - 19: em_gss_rod:*
 - 20: em_gss_vrt:*
 - 21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	850.6	51.7	972	10	AL523335
C 2	733.6	44.6	1022	10	AL523336
C 3	576.6	35.1	778	10	AL529154
C 4	567.6	34.5	642	10	AL563357
C 5	536.6	32.6	638	11	BG703339
C 6	496.4	30.2	570	11	BI341271
C 7	463.6	28.2	487	10	AW173250
C 8	436.8	26.6	578	10	AW663941
C 9	423.2	25.7	468	11	BF775076
C 10	399.4	24.3	662	11	BI394002
C 11	364.8	22.2	371	10	AA349847
C 12	278.8	17.0	536	10	AW965645

13	270	16.4	377	10	AW345918
14	265.4	16.1	402	10	AA350545
15	255.4	15.5	258	11	F08438
C 16	251.8	15.3	450	11	BI296940
17	246.4	15.0	299	11	T10079
18	244.8	14.9	289	10	AW418091
C 19	229.8	14.0	420	11	BF412623
C 20	229.2	13.9	355	10	A1372624
C 21	228.4	13.9	366	11	T08271
C 22	226.6	13.8	418	10	BE097508
C 23	226	13.7	941	10	AU067579
C 24	164.8	10.0	742	10	AU220193
C 25	144.2	8.8	308	11	EG056004
C 26	142.6	8.7	360	10	BE656418
27	138	8.4	441	11	T08675
C 28	134	8.2	328	11	BI134483
C 29	124.6	7.6	686	10	BE590388
C 30	122.4	7.4	489	13	AZ979001
C 31	104.6	6.4	996	13	CNS01UGB
C 32	104.6	6.4	1047	13	CNS05LV9
C 33	100.6	6.1	300	11	C57850
C 34	92.8	5.6	253	10	BB586972
C 35	92.8	5.6	868	13	CNS04275
C 36	91.2	5.5	1417	12	AK003981
C 37	90	5.5	275	10	BB593243
C 38	87.8	5.3	624	10	AU207190
C 39	86	5.2	563	11	BF641087
C 40	84	5.1	584	11	BI367271
C 41	82.2	5.0	208	10	BB590911
C 42	80.4	4.9	711	11	BF277036
C 43	79.8	4.9	377	11	C68924
C 44	77	4.7	640	10	AU056191
C 45	76.4	4.6	1047	13	CNS05PC2

ALIGNMENTS

RESULT 1
AL523335/c

LOCUS
DEFINITION

AL523335 LTI_NFL003_NBC3 Homo sapiens cDNA clone CS0DC001YN02 3

prime, mRNA sequence.

AL523335

AL523335.1 GI:12786828

EST.

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 972)

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

Contact: Genoscope

Genoscope - Centre National de Sequencage

Bp 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

Location/Qualifiers

1..972

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CS0DC001YN02"

/clone_lib="LTI_NFL003_NBC3"

/sex="male"

/tissue_type="neuroblastoma cells"

/lab_host="DH10B"

/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA

was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-stranded cDNA was digested with NotI and

cloned into the NotI and EcoRV sites of the pCMVSPORT 6

vector. Library was normalized. Library was constructed

by Life Technologies. Contact : Feng Liang Life

Tue Mar 12 15:54:38 2002

```

Db 590 CAGGTGGCATTCTGACCTCGGTGGTCTTTGTAGGCATGATGTCAGCTCCAGCTCTCTGG 649
Qy 421 ggaatattctcagaccaglacgacgagcaagaaacagggctgaagatcagcgtgctggact 480
Db 650 GGAATATCTCAGACCAGTACGCGAGGAAACAGGGCTGAAGATCAGCGTCTGGACT 709
Qy 481 ctgtactatggratecttagtcatttgcgcccgtgtatagctgatactgctgctccgg 540
Db 710 CTGTACTATGGCATCTTGTAGTCATTTGGCCCGGTGTATAGCTGGATCTGTGCTCCGG 769
Qy 541 ggcctagtgggcttcgggacagagagtgctcccccagtcggtgacgctgatatcccgatgc 600
Db 770 GGCTGTGGCTTCGGGATCGGAGAGTTCCCCAGTCGCTGACGCTGTATGCCAGTTTC 829
Qy 601 ctteccatgaagccagagctaaatgatttctgctgattgaggtattctggccatcggg 660
Db 830 CTTCCTCATGAWGCCAGAGCTAAATGATTTTGTCTGATTGAGGTATTCCTGGGCCATCGGG 889
Qy 561 acagttctcagagtcctcctggtctggttctgctgattgctgacgctggtggtggtgctg 720
Db 890 ACAGTGTTCGAGGTCTCTGCTGCTGTTGCTGCTGATGCCAGCCTGGGCTGGCTGGCTG 949
Qy 721 ctcatctctcagctgctccgctctctctctctctctctctctctctctctctctctctg 780
Db 950 CTCTAT-CTCTCAGCTGTC--GCTCCCTCTCTTCCGCTGCTGTTCTTGGCTGC--TGA 1004
Qy 781 agtcaaggtatgatg 798
Db 1005 AATGCAAGGATGATGTG 1022

RESULT 3
AL529154 778 bp mRNA EST 13-FEB-2001
LOCUS AL529154 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DD004YD17 5
DEFINITION prime, mRNA sequence.
ACCESSION AL529154
VERSION AL529154.1 GI:12792647
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 778)
AUTHORS Li,W.B., Gruber,C., Jessee,J., Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1..778
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DD004YD17"
/clone_lib="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 150 a 193 c 254 g 166 t 15 others
ORIGIN

```

```

Query Match 35.1% Score 576.6; DB 10; Length 778;
Best Local Similarity 97.3%; Pred. No. 2e-138;
Matches 581; Conservative 13; Mismatches 2; Indels 1; Gaps 1;
Qy 1 atgagaggaggacttattccagctaagcagctgcccgg-ttgtgaaattccgtccacagg 59
Db 182 ATGAGGAGGAGCTTATTCCAGCTAAGCAGCCCGGNTTGTGAANTCCGTCGCACAGG 241
Qy 60 cgaagtcgaagtcagagagcagacagcgttccaggagagcatgaagtcagagattgaagg 119
Db 242 CGAGAGTCAAGGTCAGAGGAGCAGACAGCGCTTCAGGAGAGCATGAAGTCCAGATTGAAG 301
Qy 120 ggtccacatggccctagaggctgtggagctggatgggagcagctgacccaagagtt 179
Db 302 GGTCCACGCTGGGCCCTAGAGGCTGTGCAGCTGGATGGGCGAGCTGTGCCCAAGAGTT 361
Qy 180 tgcgaatccccccagctgatactttcatggtgaaagatcagtggaagccattggctttgg 239
Db 362 TGCCAAATCCCAACCGATGATCTTTCATGGTGAAGATGCAGTGAAGCATTTGGCTTTGG 421
Qy 240 aaaaattccagtggaagctgtgttctcactggcttggcttggatggatgcatgga 299
Db 422 AAAATTTTCAGTGGAAAGCTGTCTGTCTCACCTGGCTTGGCTGGATGGCTGCCATGA 481
Qy 300 gatgatgattcagcatcctcagcaccagctgacccagctgcatcagtgagagctcccaagctg 359
Db 482 GATGATGATCTCTCAGCATCCTGGCACCCAGAGCTGCTTCCGAGTGGAGGCTCCCAAGCTG 541
Qy 360 gcaggtggcattgctgacccctggtgttctttaggcagctgctccagctccacgctg 419
Db 542 GCAGGTGGCATTCCTGAMCTCGTGTGCTTTTATAGCATGANGTCCAGCTCCACGCTCTG 601
Qy 420 gggaaatatctcagacacagctacggcaggaagaaacagagctgaagatcagctgtggac 479
Db 602 GGGAAAWATYTCACAGCAGTACGGCAGGAAACARGGCTGAAGATCAGGCTGCTGTGGAC 661
Qy 480 tctgactatggcattccttagtcatttgcgccgtgtatagctgatactggtgctccg 539
Db 662 TCTKWTATGGCATCTTATGTCATTTGCGCCCGTKTWTAGCTGGATCCTGTGCTCGG 721
Qy 540 gggcctggtgggcttcgggagcagagagagagagagagagagagagagagagagagag 596
Db 722 GGGCCTGCTGGGCTTCGGGATCGGAGGAGTTCGCCAGTCGGTACGCTKTTTGCCGA 778

RESULT 4
AL563357 642 bp mRNA EST 16-FEB-2001
LOCUS AL563357 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DD004YD17 3
DEFINITION prime, mRNA sequence.
ACCESSION AL563357
VERSION AL563357.1 GI:12912671
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 642)
AUTHORS Li,W.B., Gruber,C., Jessee,J., Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1..642
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DD004YD17"
/clone_lib="LTI_NFL001_NBC4"
/sex="male"

```

/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and Eco RV sites of the pCMVSPORT 6 cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"
BASE COUNT 149 a 132 c 170 g 117 t 14 others
ORIGIN

Query Match 34.5%; Score 567.6; DB 10; Length 642;
Best Local Similarity 95.9%; Pred. No. 3.9e-136;
Matches 584; Conservative 2; Mismatches 22; Indels 1; Gaps 1;
QY 1037 atgtctgagcattccagtcggaagagctgttagaggaacaaatcagctgacg 1096
DB 642 ATTTTCGGCGTCTCCAGTCGAAGAAGGCTGAAGAGGCAAAATGAGCTGGCCTGCG 583
QY 1097 agtactgagtgagagattacatgactgtgtgtgacacacctctctgagttccag 1156
DB 582 AGTAC TTGAGTGAGAGGATWACATGGACTTGCTGTGGACCACTCTCTCAGTTTCCAG 523
QY 1157 gtgtctgtgactctgtggtattattgacccgctggggcgcaagaagacattgcccctgt 1216
DB 522 GTGTCTGTGACTCTGTGGATTATTGACCGCTGTGGCGCAAGAAGACCATGGCCCTGG 463
QY 1217 gcttgcattctctctctgagctcctgctgttttctgtgttggaag-aaattgtg 1275
DB 462 GCTTTGTGATCTCTCTTCNGCACCNCNGCTGTTTNNCTGTGTGGAGAAATGNG 403
QY 1276 ctcaactgttactctctcattgcaagagcgtttatttctgagggctttcaagcgccatatt 1335
DB 402 CTCACCTCTGTAACCTCTCATGCAAGAGCGTTTATTTCTGGAGGCTTTCAAGCGGCATAT 343
QY 1336 gtttacacactgaggtctacccacgcaacgagcgccctgctggcgacactgcagc 1395
DB 342 GTTNACACACCTGAGGTCTACCCCAACGCAACCGCGGCCCTCGGCTGGCACCTGCAGC 283
QY 1396 ggcattgcaagagtggtgtctctcatcctcctgttcctgcccagggtgagtggaattcc 1455
DB 282 GGCATGCAAGAGNGGGGCTCTCANTACTCCGTTTCATCGCCACAGTGTGTAATCC 223
QY 1456 tctgtgtacttactctgagcagtttacagtgctgctgctcctctgctgcccgcctcc 1515
DB 222 TCTGTGTACTGACTCTGCGAGTTAACAGTGGCNGCTGCTCTGCTGCCCTGGCCCTCC 163
QY 1516 tgccttttgcctattgagacaaagcgagagactgcagagagtcacagccacggagtg 1575
DB 162 TGTCTTTGCCCATTTGAGACCAAGGCCGAGNACTGCAGAGTCCAGGCCACCGGAGTGG 103
QY 1576 ggcagagagatggtcgccgaggaatgcacggtgcaggtgtttaccaggtcgacactctg 1635
DB 102 GGCAGAGATGTCGGCGGAGGAATGCACGGTGCAGGTTTACCAGGTCGAGCTCTGGC 43
QY 1636 tctcagaa 1644
DB 42 TCTCAGAA 34

RESULT 5
BG703339
LOCUS BG703339 638 bp mRNA EST 07-MAY-2001
DEFINITION 602865286F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4817829 5', mRNA sequence.
ACCESSION BG703339
VERSION BG703339.1 GI:13975571
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 638)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLNL0718 row: 1 column: 22
High quality sequence stop: 638.
Location/Qualifiers
1..638
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4817829"
/clone_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTWN-3', size-selected for average insert size 2.5 kb and normalized to 500 ng. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 133 a 149 c 212 g 144 t
ORIGIN

Query Match 32.6%; Score 536.6; DB 11; Length 638;
Best Local Similarity 98.4%; Pred. No. 4.1e-128;
Matches 605; Conservative 0; Mismatches 4; Indels 6; Gaps 6;
QY 1 atgagagagagcttattccagtaaggcagctgcgcttggaattccctgcacagc 60
DB 25 ATGGAGGAGGACTTATTCACAGCTAAGCGAGCTGCCGTTGTGAATTCCTCGCACAGC 84
QY 61 gagagtcaaggtcagagacacacgcttcagagacatgaagtcacagattgaagg 120
DB 85 GAGAGTCAAGGTCAAGGACGACGACGCTTCAGGAGAGCATGAATCCAGATTGAAGG 144
QY 121 gtccacgtggccttagagctgtggagctgagatggggcagctgtgcccagagttt 180
DB 145 GTCCACGTGGGCTAGAGGCTGTGGAGCTGGATATGGGCGAGCTGTGCCAAGAGTTT 204
QY 181 gccaatccacagatgacttctcatggtggaagatgcagtggaagccattggcttgg 240
DB 205 GCCAATCCACATGATGATACATTTTCATGTTGGAAGATGCAGTGAAGCCATTGGCTTGA 264
QY 241 a-aatttcagtggaagctgtcttctcactggctggctggatggatgcatcagga 299
DB 265 ACAATTTTCAGTGAAGCTGTGTTCTCACTGGCTTGGCTTGGATGGCTGATGCATGGA 324
QY 300 gatgatgctcctcagcattcctggcaccacagctgcattgcagtgaggctcccaagctg 359
DB 325 GATGATGATCTTCACATCTCGGCACCAAGCTGATTTGCCAGTGGAGGCTCCCAAGCTG 384
QY 360 gcagtggtcattgtgcacctcgtgtgttcttgatggcatgatgtccagctccagctctg 419
DB 385 GCAGGTGGCATTTGCTGACCTCGGTGGTCTTTTGTAGGCATGATGTCCAGCTCACGCTCTG 444

```
QY 420 ggaataatctcagaccagtcagca-ggaataacaggctgaagatcagcgtgctglgga 478
|||||
Db 445 GGAATAATCTCAGACCAGTACGGCAGGAAACAGGCTGACGATCAGCGTGTGTGGA 504
|||||
QY 479 ctctgla-ctatggatccttagtcatttggcccgctgtgtatagctggatcctggctc 537
|||||
Db 505 CTCTGTACCTATGGCATCTTACTAGTCATATGG-CCGTGTATAGTGGATCCGTGTGCTC 563
|||||
QY 538 cgggacctggtgggttcaggat-cggaggacttcccc-aggcgtgacgctgtgtatgcg 595
|||||
Db 564 CGGGGCTGTGGGCTTCCGGGATACGGAGGAGTTCCTCCCAAGTCCGTGACGCTGTATGCCG 623
|||||
QY 596 agtctctcccatga 610
|||||
Db 624 AGTACCTTCCCATGA 638
|||||

RESULT 6
BI341271 570 bp mRNA EST 30-JUL-2001
LOCUS 368657 MARC 2PTG Sus scrofa cDNA 5', mRNA sequence.
DEFINITION BI341271
ACCESSION BI341271
VERSION BI341271.1 GI:15034560
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 570)
Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W.
and Keefe, J.W.
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCCAGTCACGACG
Plate: 108 row: D column: 4
Seq primer: ATTAGGTGACACTATAG.
Location/Qualifiers
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 2PTG"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT 114 a 147 c 156 g 153 t
ORIGIN
source

Query Match 30.2%; Score 496.4; DB 11; Length 570;
Best Local Similarity 91.9%; Pred. No. 9.8e-118;
Matches 524; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 619 gctaaatgatttctgattgaggtattctgggccatcgggacagtggttcgaggtcgc 678
|||||
Db 1 GCTAAGTGCATTTTCTGATTCAGGTGTCTCGGCCATCGGACGGTGTTCGAGGTGTCATC 60
|||||
QY 679 ctggctgtgttcgatgccagcctgggctggcggttggtgctcatctcagctc 738
|||||
```

```
Db 61 CTGGCTGTGTTTGTGATGCCAGCCTGGGCTGGCGTTGGCTGCTCTCTTTTCAGCTGTC 120
QY 739 ccgctcctcctcttgcgtgctgtgttctggctgctgaagctgaagtgatgatg 798
|||||
Db 121 CCGCTCCTCTCTTCCGTGCTCCTGTTTGGCTGGCAGAGATGCAAGATATGATGTG 180
QY 799 ctgtcagggaaacaggaaagcaatcgccacctaaagagatagcaactgaaacqga 858
|||||
Db 181 CTCCTCTGAAACACAGGAAAGCCATTGCCACCTTTAAAGAGATAGCCACGGAACCGA 240
QY 859 gctccatcgccgtgggaaactcatctccagacaggagaccgagcgaataaggg 918
|||||
Db 241 GCCCCATCCCTCTGGGAAGCTCATCATCTCCAGACAGGAAGACCGGCAAAATGAGG 300
QY 919 gacctttcacacccctttagatggacaacttggctgtgtgttttatgttttc 978
|||||
Db 301 GACCTTTTTCACACCCCATTTTAGATGGACAACCTTTGTCGTGTGTATATGTTTTC 360
QY 979 aatgcattcttactacgggttagtttactcacccagaaactctccagcgaggagat 1038
|||||
Db 361 AATGGGTTTCTTACTATAGGACTAGTTCTGCTCACCACTGAGCTCTCCAGGCTGGAGAT 420
QY 1039 gtctgcgcatctccagtcggaagaggctgtagaggcaaaatcagcctgctgcgag 1098
|||||
Db 421 GTCTGCAGCATCTCCAGCGGGAAGGCGCTAGAGGCAAAATGCAGCTGGCCTGCCGAG 480
QY 1099 tacctgagtgagagattacatgactgtgtgtgacccctctctgagtttccaggt 1158
|||||
Db 481 TACCTGAGTAAGAGAGATTACATGACACCTGCTGTGACCACGCTCTCTCAGTTTCCAGGT 540
QY 1159 gtctgtgactctgtggattattgaccgc 1188
|||||
Db 541 GTCTGTGACTCTCTGGATTATCGACCGC 570

RESULT 7
AW173250/c 487 bp mRNA EST 16-NOV-1999
LOCUS xj85b11.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:2663997 3' similar to TR:Q9Z217 Q9Z217 SV2 RELATED PROTEIN.
; mRNA sequence.
ACCESSION AW173250
VERSION AW173250.1 GI:6439198
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 487)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapb@remail.nih.gov
This clone is available royalty-free through LNL : contact the
TRACE Consortium (info@image.lnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 1.
Location/Qualifiers
1. .487
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2663997"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19w, testis NHT, and B-cell
NCI CGAP GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
```


tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same libraries. The pools consisted of 1 M.A.C. clones 297480-302087, 682632-687239, 726408-726711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo. ^a Others 126

Query Match	28.2%	Score 463.6	DB 10	Length 487
Best Local Similarity	96.9%	Pred. No. 2.8e-109		
Matches 472	Conservative 0	Mismatches 15	Indels 0	Gaps 0
Qy 1156	gggtgctctgtgactctgtgattattgaacgcctcggggcgcaagaagaccatggccctg	1215		
Db 487	GGTGTCAATGTGACTCTGTGGATTAATGACCGCATGGGGCGCAAGAAGACCAATGGCCNTG	428		
Qy 1216	tactgtgcatctctctctctcgcagcctcctcgtgtttatctgtgttggaaagaaatg	1275		
Db 427	TGCTTGCTCAATCTTCTCTCTGTGAGCCTCGTGTCTGTTTATCTGTGTGAGAGATGTG	368		
Qy 1276	ctcaectctgttactcttcattgcaagagcgtttattctgagagccttccaagcgcata	1335		
Db 367	CTCACATCTCTTACTTCTTATTCGAAGACGCTTATTAATGAGAGCCTTCAACCGGCATAT	308		
Qy 1336	gtttacacacctgaggtctaccacacggaacgcggccctcgccctggtggcacctgcac	1395		
Db 307	GTTTACACACGTGAGGTCTACCCCAACGGCAACGGGGCCCTCGGCCCTGGGCACCTGCAGC	248		
Qy 1396	ggcatggcaagagtggtgtctctcaactcgttctatcgccacaggtgatctgggaatcc	1455		
Db 247	GGCATGGCAAGAGTGGGTGCTCTCATCACTCCGTTCATCGCCACAGGTGATGCTGGAAATCC	188		
Qy 1456	tctgtgtacctgactctggcagtttacagtggtgtcgtcctctcgtcgtcctcgtccct	1515		
Db 187	TCGTGTACCTGACTCTGGCAGTTTTACAGTGGCTGCTGCCCTTCGTGCTCGCCCTGGCCCTCC	128		
Qy 1516	tgtttttgcccatgagaccaagcgagagactgcaggagtcacagccaccgcggagctg	1575		
Db 127	TGCTTTTGGCCATTGAGACCAAGGCCGAGGACTGCAGGAGTCCAGCCACGGGAGTGG	68		
Qy 1576	ggccaggagatggtcggccgaggaatgacaggtgcaggtgttacccaggtcaactctggc	1635		
Db 67	GCCCAGGAGATGGTCGGGCCAAGGAATGCACGGTTCAGGTTCCAGGTTCCAGGTCGAATCTGGC	8		
Qy 1636	tctcagg 1642			
Db 7	TCTCAGG 1			
RESULT 8				
AW663941/c				
LOCUS	578 bp	mRNA	EST	06-APR-2000
DEFINITION	h173h10.x1 Soares.NFL.T_GBC.S1 Homo sapiens cDNA clone			
	IMAGE:2977987.3, similar to TR:Q92217 Q92217 SV2 RELATED PROTEIN.			
	.. mRNA sequence.			

AW663941
 VERSION
 AW663941.1 GI:7456480
 EST.
 KEYWORDS
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 578)
 REFERENCE
 NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 AUTHORS
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 TITLE
 Tumor Gene Index
 Unpublished (1997)
 JOURNAL
 Contact: Robert Strausberg, Ph.D.
 COMMENT
 Email: cgaps-r@mail.nih.gov
 This clone is available royalty-free through LNL; contact the
nlncgenetics@lncfimage.lnl.gov for further information.

```

Seq primer: -40UP from Gibco
High quality sequence stop: 464.
Location/Qualifiers
1..578
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2977987"
/clone_lib="Soares_NFL_T_GBC_S1"
/clone_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site: 1: Not I; Site: 2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI_CGAP_GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Ronaldo."
137 a 170 c 159 g 111 t 1 others

```

Query Match	26.68;	Score 436.8;	DB 10;	Length 578;
Best Local Similarity	99.18;	Pred. No. 2.6e-102;		
Matches 449;	Conservative 0;	Mismatches 3;	Indels 1;	Gaps 1;
QY	1193	ggcgcgaagaagaccat-gggccctgtgctttgttcattctctctctctgcagccctctgcgtc	1251	
Db	578	GGCGCAAAAAGACCATGGGCCCTGTGCTTTGTTCATCTCTCTCTGACGCCCTCTGCTG	519	
QY	1252	tttatctgtgttggagaanaatgctcactctgttaccttcatcttcaagagcggtttat	1311	
Db	518	TTTATCTGTGTTGGGAAGANATGTGCTCACTCTGTTTACTTCTTCATTTGCAAGCGTTTATT	459	
QY	1312	tctggagcgcttcaagcgccatattgttaccacactgaggtctaccctcagcgcaacgcg	1371	
Db	458	TCFGGAGGCTTTCAGCGGCATATCTTTACACACCTGAGGTCTACCCACGGCAACGCGG	399	
QY	1372	gccttcggcctggggcaactgcagcgccatggaagagtggtgtctctcatcactccgttc	1431	
Db	398	GCCTCTGGCTGGGCAACCTGCAGCGGCATGGCAAGAGTGGGTGCTCTCATCATCTCCGTC	339	
QY	1432	atgcgccaggtgaatcctggaaacctctgtacctgactctggcagtttacagtggtctgc	1491	
Db	338	ATGCCCCAGGTGATGCTTGAATCTCTGTGTACCTTGACTCTGGCAGTTTACAGTGGCTGC	279	
QY	1492	tgcctctggctgcccctggcctctctgttlttgcccatctgagacccaagagcgaggaactg	1551	
Db	278	TGCTCTCTGGCTGCCCTGGCTCTCTGCTTTTGGCCATTGAGACCAAAAGCGCCGAGGACTG	219	
QY	1552	caggagttccagccacccgggaagtggggccaggaatggtctgcgcaggaagaatcacggttgc	1611	
Db	218	CAGGAGTCCAGCCACCCGGGAGTGGGGCCAGGAGATGGTCCGCCGAGGAATCACCGGTGCA	159	
QY	1612	ggtgttaccaggttcgaactctgctctcaggaa	1644	
Db	158	CTCTTCTACCGTTCGAACCTGCTGCTCTCAGGAA	126	

RESULT	9				EST	25-APR-2001
LOCUS	Bf775076		468 bp	mrna		
DEFINITION	255128	MARC 3BOV Bos taurus	cdna 5'			EST
ACCESSION	Bf775076					
VERSION	Bf775076.1	GI:12122976				
KEYWORDS	EST.					
SOURCE	COW.					
ORGANISM	Bos taurus					
	Eukaryota;	Metazoa;	Chordata;	Vertebrata;	Euteleostomi;	
	Mammalia;	Eutheria;	Cetartiodactyla;	Ruminantia;	Pecora;	Bovidae

Bovidae; Bovinae; Bos.
 1 (bases 1 to 468)
 Smith,T.P.L., Grosche,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
 Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,
 G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
 Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
 Keele,J.W.
TITLE
 Sequence evaluation of four pooled-tissue normalized bovine cDNA
 libraries and construction of a gene index for cattle
 Genome Res. 11 (4), 626-630 (2001)

JOURNAL
MEDLINE
COMMENT
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCACAGTCACGACG
Plate: 86 row: E column: 5
Seq primer: ATTTAGTGACTACTATAG.
Location/Qualifiers
1 . 468
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 3BOV"
/lisuse_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6;
Library made from marrow, alveolar macrophage, ovary, fetal semitenodonus muscle, and fetal longissimus muscle."
BASE COUNT 87 A 126 C 139 G 116 T

FEATURES
source
1..468
/organism="Gallus gallus"
/strain="Commercial broiler chicken"
/db_xref="taxon:9031"
/clone_lib="pk013.c4"
/dev_stage="Embryonic (d12,d14,d19); post-hatch (w1,w3,w5,w7,w9)"
/lab_host="E. Coli EM DH10B"
/note="Vector: pCMVSPORT6; Library made from equivalent pools of total RNA isolated from each tissue at different ages. Single pass sequencing from 5'-end"

Query Match Best Local Similarity Matches	24.3% 79.0%; Pred. No. 1.3e+92;	Score 399.4; DB 11; Length 662; Conservative 0; Mismatches 126; Indels 0; Gaps 0;
---	------------------------------------	--

QY 487 tatggcatcttagtcatttgcgcccgtgatagatcgacttgccgggccgtg 546		
Df 1 TAGGGCATCTCAGTCGTGTTGCCCGGTTACAGCTGGATCCGTGCCTCGGGGCCCTG 60		
<hr/>		
QY 547 gtgggtttoggatcaggagagtccccagtcggtaacgttatgccgaattccctccc 606		
Df 61 GTGGGTTTCGGATTGGAGGGTCCCCAGTCGGTCACACTGATGCTGAGTTCCTCCC 120		
<hr/>		
QY 607 ataaaccagagaataaatgtatttgtgattgagtgatttggccatcgaggacagt 666		
Df 121 ATNAAGCACAGACTAAATGTAATTTGCTGATTGAGTGTCTTGCGCCATCGGACAGTG 180		
<hr/>		
QY 667 ttcaaggtctcctgcgtggttcgtgatgccagctgggctggcgtggcgtcac 726		
Df 181 TTCAGAGTCTCTGGCTGTGTTGTGATGCCAGGCTGGCTGGCTGGCTGGCTGGCTCA 240		
<hr/>		
QY 727 ctctcagctcctcctcctcttgcctgctgctgtgttcttgcctgctaagtcca 786		
Df 241 CTCTCAGCTGTCCACTCTCTCTCTTTCGGTCTGTGTTTTTGGCTGCCGAGAGTGG 300		
<hr/>		
QY 787 aggtatgatgtgcttcaggaaaaccagaaaaaacgaatccacccttaaaggatatgca 846		
Df 301 AGGTATGATGTGTGTCGGGAACAAGAACGAGGCCATGCCACCTAAGAAGATAAGCC 360		
<hr/>		
QY 847 actgaacaagagctccatcgccgtggtgggaacctcatctccagacagggaagacca 906		
Df 361 ACAGAAAACGAGCTCCCTCTCTGCGGAAGAACTCATCATTTCCAGACAGGAAGACCCA 920		
<hr/>		
QY 907 ggcaaataggggaacttttacacocccatttagatggacaactttg 954		

Qy	985	ttctcttactacgggttagttctactaccacagaaactcttccaggcagagatgtctgc	1044
Db	1	TTCTCTTACTACGGGTAGTTCTACTCACCAGAACTCTTCCAGGAGGAGATGTCTGC	60
Qy	1045	ggcatctccagtcggaagaagagctctagaggcaaaaatgcagacctggcctgcgagctaccctg	1104
Db	61	GGCATCTCCAGTCGGAAGAAGAGCTGTAGAGGCGAAATCGACGCTTGGCCCTGCGAGTACTCG	120
Qy	1105	agtgaagagattacatggaactctctggagcacccctctctgagtttccagggtgcctt	1164
Db	121	AGTNAGGNGGATTACATGAGCTGCTGTGGACCACTCTCTGAGTTTCCAGGTGTCTCT	180
Qy	1165	gtgactctgtggattattgacgcctctggggcgcaagaagaccatggccctctgcttctgc	1224
Db	181	GTGACTCTNTGGATTATTGACCGCCTGGGGCGCAAGAAGACCATGGCCCTGTGCTTTGTC	240
Qy	1225	atctctctctctgcagcctcctgctgtttatctatgttgtgaagaatgtgctcactctg	1284
Db	241	ATCTTCTCCCTCTGCAAGCCTCCCTGCTGTATTCTGTGTGTGGAAGAAATGTGCTCACTCTG	300
Qy	1285	ttactcttcattcgaagacggtttattctctggaggcttttccaggcgccatatgtttacaca	1344
Db	301	TTACTCTTCATTGCAAGAGGCTTTATTCTTGGAGGCTTTTCAAGCGGCATATGTTTACAAA	360
Qy	1345	cctgaggtcta	1355
Db	361	ACTGAGGTCTA	371

RESULT	12
LOCUS	AW965645 536 bp mRNA EST 01-JUN-2000
DEFINITION	F-137771.p MAGE rescuesques, MAGI Homo sapiens cDNA, mRNA sequence.
ACCESSION	AW965645
VERSION	AW965645.1 GI:8155481
KEYWORDS	EST..
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Karyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 536)
AUTHORS	Hegde,P., Qi,R., Abernathy,K., Bharap,S., Gaspard,R., Gay,C., Holt, I.E., Sauced,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and Quackenbush,J.
TITLE	Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray
JOURNAL	Unpublished (2000)
COMMENT	Contact: John Quackenbush

```

seq prim
FEATURES             location/Qualifiers
    source            1..536
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone_lib="MAGE_rescueses, MAGI"
                        /note="vector: pBluescriptSkm"
BASE COUNT          95 a 144 c 174 g 123 t
ORIGIN

```

```

Query Match      17.0%; Score 278.8; DB 10; Length 536;
Best Local Similarity 99.3%; Pred. No. 1.9e-61;
Matches 280; Conservative 2; Indels 0; Gaps 0;

```


/tissue_type="total brain"
/dev_stage="3 months old"
/note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
Site_2: NotI; sex=female; dev_stage=3 months old;
isolate-muscular atrophy patient; tissue_type=total brain
; total mRNA was oligo-(dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the
lafmid BA vector. Clone library from B.Souares, psychiatry
Dept. Columbia University, USA. Normalization_method:
Bento Soares, P.N.A.S in press"
BASE COUNT 57 a 55 c 63 t 1 others
ORIGIN

Query Match 15.5% Score 255.4; DB 11; Length 258;
Best Local Similarity 99.2%; Pred. No. 1.6e-55;
Matches 256; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 80 acgacacgcttcagagagcatgaagtcacagattgaagggtccacgtggcctagagg 139
DB 1 ACGACACGGCTTCAGGAGAGCATGAAGTCCAGATTGAAGGGGTCCACGTGGCCCTAGAGG 60
QY 140 ctgtgagctggatgagggcagctgtgcccagaggtttgcccattcccaccgatgata 199
DB 61 CTGTGAGCTGGATGATGGGGCAGCTGTCCCAAGAGCTTTCACCAATCCCACTGATGATA 120
QY 200 cttctatggtggaagatgcagtggaagccattggtcttggaaaaatttcagtgaaagctg 259
DB 121 CTTTCATGTGTGGAAGATGCAGTGAAGCCATTGGCTTTTGGAAAAATTCAGTGGAGCTGT 180
QY 260 ctgttctcactggtctggttgatggtgctgacatggagatgatgatccctcagatccc 319
DB 181 CTGTTCTCACTGGCTTGGCTTGGATGGCTGATGCCATGGNGATGATGATCCTCAGCATCC 240
QY 320 tggcaccacagctgcatt 337
DB 241 TGGCACCACAGCTGCATT 258

Search completed: March 8, 2002, 00:48:17
Job time: 4640 sec

/db_xref="taxon:9606"
/clone_lib="infant brain"
/sex="female"
/dev_stage="infant"
/note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
Site_2: NotI"
BASE COUNT 72 a 108 c 129 g 85 t 8 others
ORIGIN

Query Match 16.1%; Score 265.4; DB 10; Length 402;
Best Local Similarity 97.8%; Pred. No. 5e-58;
Matches 269; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1370 gggccctcgtggcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1429
DB 23 GGGGCCCTCGGCTGGCCCTGCGCCGATGGCAGAGTGGGTGCTCTCATCCTCCTGCT 82
QY 1430 tcatgcccagtgatgctggaatcctctgtgtacatgactctggtgagcagtgagct 1489
DB 83 TCATGCCCCAGGTGATGCTGGAATCTCTGTGTACTGTGCTGCTGCTGCTGCTGCTGCT 142
QY 1490 gctgctcctgagcctgagcctcctgctcttggccattgagaccaaagcgagagac 1549
DB 143 GCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 202
QY 1550 tgcagagtcacacccagcagcagcagcagcagcagcagcagcagcagcagcagcag 1609
DB 203 TGCAGGAGTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 262
QY 1610 caggtgtaccagtcgaactcgtgctcagga 1644
DB 263 CAGGTGTTACCAAGTTCGAACCTCTGCTCTCAGGAA 297

RESULT 15
F08438 258 bp mRNA EST 22-FEB-1995
LOCUS HSC2UB081 normalized infant brain cDNA Homo sapiens cDNA clone
c-zub08, mRNA sequence.
F08438
F08438.1 GI:677005
EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 258)
AUTHORS Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes
M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F.,
Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,f.,
Sebastiani-Kabaktchis,C. and Tessier,A.
TITLE IMAGE: molecular integration of the analysis of the human genome
and its expression
JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
MEDLINE 95277534
COMMENT Contact: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1,rue de l'Internationale, Bp60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr
Single read.
Genexpress_library_idt: C; Genexpress_sequence_idt: y1c-zub08
Seq primer: (-21)M13-universal.
Location/Qualifiers
FEATURES
source
1..258
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-zub08"
/clone_lib="normalized infant brain cDNA"
/sex="Female"

• OM of: US-09-911-667A-2 to: GenEmbl.* out_format : pfs
Date: Mar 8, 2002 3:34 AM
About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-O=/cgn2.1/USPTO_spool/US09911667/runat_07032002_130318_29538/app_query.fasta_1.615
-DB=GenEmbl -OFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINWATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US09911667.ecgnl.1.6277 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-911-667A-2
Query length: 548
Database: GenEmbl.*
Database sequences: 1472140
Database length: 341344837
Search time (sec): 1970.640000

score_list:

Sequence	Strd	Orig	zScore	Escore	Len	Documentation
gb_ro:AF060173	+	2749.00	3428.33	1.1e-182	2421	! AF060173 Rattus norvegicus SV2
gb_in:AC004642	+	984.00	1184.24	1.1e-57	18432	! AC004642 Drosophila melanoga
gb_hg:AC020509	-	984.00	1183.68	1.2e-57	157851	! AC020509 Drosophila melanoga
gb_in:AE003462	-	984.00	1177.90	2.5e-57	300542	! AE003462 Drosophila melanoga
gb_pr:HS0802697	+	734.00	909.82	2.1e-42	2104	! AL359592 Homo sapiens mRNA; ch
gb_pl:AB026645	+	723.50	863.92	7.6e-40	81020	! AB026645 Arabidopsis thaliana
em_pl:AC024128	+	723.50	863.92	7.6e-40	81020	! AC024128 Arabidopsis thaliana
gb_in:AE001863	+	644.50	757.07	6.8e-34	198677	! AE001863 Deinococcus radiodu
gb_in:CEK01F9	-	640.50	772.76	9.1e-35	19834	! Z22175 Caenorhabditis elegans
gb_hg:H0702G05	+	572.00	669.94	4.9e-29	134092	! AL442105 Oryza sativa chromo
gb_ba:AE008141	-	568.50	687.78	4.9e-30	11286	! AE008141 Agrobacterium tumefa
gb_ba:AE007266	-	540.50	652.33	4.7e-28	11856	! AE007266 Shorhizobium melill
gb_ba:AE001804	+	527.50	632.50	5.9e-27	17641	! AE001804 Theriotoga maritima
gb_in:CEZK637	-	511.50	604.99	2.0e-25	40699	! Z11115 Caenorhabditis elegans
gb_pat:AX127153	+	498.50	570.52	1.7e-23	303400	! AX127153 Sequence 7069 from
gb_pat:AX123379	+	493.50	613.10	7.1e-26	1344	! AX123379 Sequence 3295 from
gb_pat:AX066931	+	493.50	612.48	7.7e-26	1440	! AX066931 Sequence 513 from
gb_ba:AB000617	+	491.00	580.71	4.5e-24	35000	! AB000617 Bacillus subtilis ge
gb_ba:BS080002	+	491.00	564.16	3.8e-23	221160	! Z99105 Bacillus subtilis com
gb_ba:AE004767	+	466.50	561.02	5.7e-23	10344	! AE004767 Pseudomonas aerugin
gb_pat:E31258	+	446.50	551.22	2.0e-22	1901	! E31258 Liver-specific organic
gb_pat:E31259	+	446.50	551.22	2.0e-22	1901	! E31259 Liver-specific organic
gb_ro:RATLSTP	+	437.50	539.92	8.9e-22	1910	! D27651 Rattus norvegicus lived
gb_ro:DB3044	+	436.00	536.97	1.2e-21	2152	! DB3044 Rattus norvegicus mRNA
gb_ro:RNOCT2R	+	435.00	535.92	1.4e-21	2106	! Y13154 Rattus norvegicus mRNA
gb_ro:RNOCTK	+	435.00	535.80	1.4e-21	2132	! X98334 R.norvegicus mRNA for
gb_ro:RNCATTRA	+	431.00	531.92	2.4e-21	1882	! X78855 R.norvegicus mRNA for
gb_pat:A8157	+	430.00	530.67	2.8e-21	1882	! A8157 Sequence 4 from Patent
gb_om:SSOCT2P	+	429.00	525.31	5.9e-21	2975	! Y09400 S.scrofa mRNA for apica
gb_pr:HS077086	+	427.00	526.98	4.5e-21	1870	! Y07086 Human organic cation tr
gb_pat:A48159	+	426.00	525.60	5.3e-21	1896	! A48159 Sequence 6 from Patent
gb_pat:A48158	+	425.50	525.03	5.7e-21	1885	! A48158 Sequence 5 from Patent
gb_ro:MMU38652	+	425.00	524.41	6.2e-21	1884	! U38652 Mus musculus transmembr
gb_pr:HSOCTL	+	425.00	524.39	6.2e-21	1888	! X98332 H.sapiens mRNA for orga
gb_ro:AF010259	+	423.50	522.60	7.8e-21	1870	! AF010259 Mus musculus organic
gb_ro:MMU6036	+	423.00	520.67	7.8e-21	2164	! A006036 Mus musculus organic
gb_pr:HSOCTK	+	423.00	520.29	1.1e-20	2257	! X98333 H.sapiens mRNA for orga
gb_ro:AF060174	+	420.00	515.19	2.0e-20	2622	! AF060174 Rattus norvegicus syn
gb_ro:RATSV2A	+	413.00	503.43	9.1e-20	3663	! L01788 Rattus norvegicus synap
gb_ro:RATSV2A	+	412.00	501.75	1.1e-19	3844	! L05435 Rattus norvegicus synap
gb_pr:AB018278	+	409.50	496.36	2.3e-19	4948	! AB018278 Homo sapiens mRNA for

gb_om:AF015958 + 409.00 501.07 1.2e-19 2732 ! AF015958 Oryctolagus cunicu
gb_pr:AB018279 + 408.50 496.26 2.3e-19 4353 ! AB018279 Homo sapiens mRNA
gb_pat:AX074148 + 407.00 500.84 1.3e-19 2121 ! AX074148 Sequence 4 from Pa
gb_ro:RATSV2BA + 406.00 494.69 2.8e-19 3660 ! L10362 Rattus norvegicus sy
seq_name: gb_ro:AF060173
seq_documentation_block: 2421 bp mRNA ROD 21-NOV-1998
LOCUS AF060173 Rattus norvegicus SV2 related protein (SVOP) mRNA, complete cds.
DEFINITION Rattus norvegicus SV2 related protein (SVOP) mRNA, complete cds.
ACCESSION AF060173
VERSION AF060173.1 GI:3901267
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 2421)
AUTHORS Janz, R., Hofmann, K. and Sudhof, T.C.
TITLE SVOP, an evolutionarily conserved synaptic vesicle protein,
suggests novel transport functions of synaptic vesicles
J. Neurosci. 18 (22), 9269-9281 (1998)
JOURNAL 99019745
MEDLINE
REFERENCE 2 (bases 1 to 2421)
AUTHORS Janz, R., Hofmann, K. and Sudhof, T.C.
TITLE Direct Submission
JOURNAL Submitted (16-APR-1998) Center for Basic Neuroscience, Department
of Molecular Genetics, HHMI, UT Southwestern Medical Center at
Dallas, 5323 Harry Hines Boulevard, Dallas, TX 75235, USA
Location/Qualifiers
1. :2421
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/tissue_type="Brain"
1. :2421
/gene="SVOP"
301. :1947
/note="SVOP"
/note="synaptic vesicle protein; similar to transport
proteins"
/codon_start=1
/product="SV2 related protein"
/protein_id="AAC78627.1"
/db_xref="GI:3901268"
/translation="MEEDLQRLPQVVKFRKTGESARSEDAAAGEHDVQIEGVVRVG
LEAVELDGAAPVKKEFANPTDDTFMVDAVEAIGFRQWKLSTGLTGLAMADAMEMM
ILSILAPQLHCWRPLPSQVALLTSVVFQGMSSSTLMGNSIDQGRKTLKISVFT
IGTVFEVLIAVFMPSLGRWLLLSAALLVAVLCFMPSPARAKILLIEFWA
TLKRIATENGAPMPLKLLISROEDRGMRLDTPHFRWTLTLLMLTFISNFAFYGL
VLLTTELFOAGVCSISSKKKAVEAKSLACEYLSKEDYMDLLMTLLSEFGVLTGLW
VIDRGRKKTALCFVIFSLCSLLIFICIGRNVLTLILFIARAFISGGFQAAYVTPE
VYPTATRALGTCSCMARVGLITPTAQVMLESSVYLTAVYSGCCLLAALASCFL
PIETGRALQESHREWGEMVGRGTNSTGVPSRNSGOE"

BASE COUNT 459 a 658 c 735 g 569 t
ORIGIN

alignment_scores:
Quality: 2749.00 Length: 548
Ratio: 5.044 Gaps: 0
Percent Similarity: 99.453 Percent Identity: 95.985

alignment_block:
US-09-911-667A-2 x AF060173 ..
Align seg 1/1 to: AF060173 from: 1 to: 2421

1 MetGluGluAspLeuPheGlnLeuArgGlnLeuProValValLysPheAr 17
|||||
301 ATGAGGAGGAGGACCTGTTCCAGCTCAGGCAGTTCGGGTGGAATCCG 350

```

17  gArgThrGlyCyluSerAlaArgSerGluAspThrAlaSerGlyGluH 34
|||||
351  CGCACAGGAGAGAGCCACGGTCAGAGGAGCGCGCTTCGGGGAAAC 400
|||||
34  IsGluValGlnIleGluGlyValHisValGlyLeuGluAlaValGluLeu 50
|||||
401  ATGATGTTTCACATTCGAGGGGTCGAGTGGGCCCTAGAAAGCTGTCGAGCTG 450
|||||
51  AspAspGlyAlaIaValProLysGluPheAlaAsnProThrAspAspTh 67
|||||
451  GATGATGAGGAGCTGTCCTCAGCGCTTCGAGGATTCGCAATCCCACTGATGACAC 500
|||||
67  rPheMetValGluAspAlaValGluAlaIleGlyPheGlyLysPheGlnT 84
|||||
501  TTTTCATGTCGAAGATGCGGTGGAAGCCATTGGGTTCGGAAGATTCGAGT 550
|||||
84  rPlsLeuSerValLeuThrGlyLeuAlaTrpMetAlaAspAlaMetGlu 100
|||||
551  GGAAGCTCTCTCTCTCAGCGCTTCGCTTGGATGGCGGAGCCCATGGAG 600
|||||
101  MetMetIleLeuSerIleLeuAlaProGlnLeuHisCysGluTrpArgLe 117
|||||
601  ATGATGATTCAGCATCTCGGCTCGGCTCAGCTGCACCTCGGAGTGGCGGACT 650
|||||
117  uProSerTrpGlnValAlaLeuThrSerValValPheValGlyMetM 134
|||||
651  CCCCAGCTGGCAGGTGGCGCTGCTGACCTTCGGTGGTCTTCATTGGTATGA 700
|||||
134  etSerSerThrLeuTrpGlyAsnIleSerAspGlnTrpGlyArgLys 150
|||||
701  TGTCCAGTTCACGCTCTGGGAAACATCTCGGATCAGTATGGCAGGAAA 750
|||||
151  ThrGlyLeuLysIleSerValLeuThrTrpLeuTrpGlyIleLeuSe 167
|||||
751  ACAGGGCTGAAGATCAGTGTGTTCTGGACCCGTACTACGGCATCTCTAG 800
|||||
167  rAlaPheAlaProValTrpSerTrpIleLeuValLeuArgGlyLeuValG 184
|||||
801  CGCTTTCCGCCAGTGTATAGCTGGATCTCGTGTCTCCGAGGCTCGGTG 850
|||||
184  lyPheGlyIleGlyValProGlnSerValThrLeuTrpAlaGluPhe 200
|||||
851  GCTTTGGGATGGAGGGGTGCCCTGAGTCGGTGACCTGTACGCCGAGTTC 900
|||||
201  LeuProMetLysAlaArgAlaLysCysIleLeuLeuIleGluValPheTr 217
|||||
901  CTCCCCATGAAGCCAGAGCCAAAGTGCATTTGCTGATTGAGGTTTCTG 950
|||||
217  pAlaIleGlyThrValPheGluValValLeuAlaValPheValMetProS 234
|||||
951  GGCCATCGGACCGCTGTTCAGGTGCTTCGCTGTGTTGTATGCCCA 1000
|||||
234  erLeuGlyTrpArgTrpLeuLeuIleLeuSerAlaValProLeuLeuLeu 250
|||||
1001  GCCTGGGCTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1050
|||||
251  PheAlaValLeuCysPheTrpLeuProGluSerAlaArgTrpAspValLe 267
|||||
1051  TTTGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1100
|||||
267  uSerGlyAsnGlnGluLysAlaIleAlaThrLeuLysArgIleAlaThrG 284
|||||
1101  GTCTGGGACCAAGGAAAGGCGATTGCTACCTTAAGCGGATCCAAACAG 1150
|||||
284  luAsnGlyAlaProMetProLeuGlyLysLeuIleIleSerArgGlnGlu 300
|||||
1151  AAAATGGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
|||||
301  AspArgGlyLysMetArgAspLeuPheThrProHisPheArgTrpThrTh 317
|||||
1201  GACCCAGGCAAAATGAGGAGCTTTTCACACCCCACTTTCTGTTGGACAAC 1250
|||||
317  rLeuLeuLeuTrpPheIleTrpPheSerAsnAlaPheSerTrpTrpGlyL 334

```

seq_name: gb_in:AC004642

seq_documentation_block:
LOCUS AC004642 148432 bp DNA INV 01-MAY-1998
DEFINITION Drosophila melanogaster DNA sequence (pIs DS00543 (D193) and DS02867 (D200)), complete sequence.
ACCESSION AC004642 AC004302 AC003551 AC003552 AC003553 AC003554 AC003555
AC003556 AC003557 AC003558 AC004308 AC003575 AC003848 AC003849
AC003576 AC003577 AC003579 AC003580 AC003581 AC003582
AC003583 AC003584 AC003585 AC003850 AC003586 AC003587 AC003588
AC003589 AC003590 AC003591 AC003592
VERSION AC004642.1 GI:3097827
KEYWORDS HTG.
SOURCE Drosophila melanogaster (Subclones in Sac from P1 clones DS00543 (D193) and DS02867 (D200)) DNA.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroilidae; Drosophilidae; Drosophila.

1 (bases 1 to 148432)
Celnikier,S.E., Aghavani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
Humast,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
Lomon,M.A., Mazda,P., Mok,M.S., Moshrefi,A.R., Moshrefi,M.,
Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,D., Snir,E.,
Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,K., Zieran,L.L.
and Kimmel,B.
Sequencing of Drosophila chromosome 2R, region 60B1-60B10
Unpublished (1997)
2 (bases 1 to 148432)
Celnikier,S.E., Aghavani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
Humast,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
Lomon,M.A., Mazda,P., Mok,M.S., Moshrefi,A.R., Moshrefi,M.,
Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,D., Snir,E.,
Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,K., Zieran,L.L.
and Kimmel,B.
Direct Submission
Submitted (01-MAY-1998) Berkeley Drosophila Genome Project, MS
64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, US
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site
(<http://fruitfly.berkeley.edu/sequence-archive.html>) or send email
to drosophila@hgsc.lbl.gov.
Library locations: 63.6 83_30.
Location/Qualifiers
1. 148432
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2R"
/map="60B1-60B10"
/clone="P1s DS00543 (D193) and DS02867 (D200)"
/note="d193 extends from bp 1 to bp 83271 and d200 extends
from bp 81802 to bp 148432."
BASE COUNT 38713 a 34662 c 35252 g 39805 t
ORIGIN

alignment_scores:
Quality: 984.00 Length: 605
Ratio: 2.589 Gaps: 19
Percent Similarity: 62.810 Percent Identity: 39.835

alignment_block:
US-09-911-667A-2 x AC004642 .. *

Align seg 1/1 to: AC004642 from: 1 to: 148432

41 ValHisValGly LeuGluAlaValGluLeuAsp...AspGlyAlaAlav 56
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
75088 ATCCACATTCGACTGAAGCGGTGCTCTATAAATCCGGATGGTAAAGTGT 75137

56 alProLysGluPhe.....AlaAsnProThrAspAsp 66
:::|||||:::|||||:::|||||:::|||||:::|||||
75138 CCGAAGAGGAATTTCCACATTTTCGGGTACCAAGTCACCTTTCCAGAT 75187

67 ThrPheMetValGluAspAlaValGluAlaIleGlyPheGlyLysPheG1 83
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
75188 ACCTACACTGTTCCAGCAGGCATCAATGCCTTCGGTTCCGATGGTTTCA 75237

83 nTrpLysLeuSerValLeuThrGlyLeuAlaTrpMetAlaAspAlaMetG 100
: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
75238 CGTCAAGCTGCTGCTTGTGTGGGGCTGCTGCTGATGTCGACGATCCATG 75287

100 LuMetMetIleLeuSerIleLeuAlaProGlnLeuHisCysGluTrpArg 116

76185	GGAGAGCTTTAGATCCCTGCTCAGCCGCGACCTGTGTACAGACCAACCATCC	76233
319	euLeuTrpPheIleTTrpPheSerAsnAlaPheSerTyrTyrGlyLeuVal	335
76235	TGTTGTGTTTCTCTGGCTGCCCTCGCCTTTTGTCTACTACGCGCTGGTG	76284
336	LeuLeuThrThrGluLeuPheGlnAlaGlyAspValcysGlyIleSerSe	352
76285	CTGTCACGACCGAGCTGCTGGTGGCC.....	76311
352	rArqLysLysAlaValaGluAlaLysCysSerLeuAlaCysGluTyrLeuS	369
76312	AGAAACAAGAGAGCATCCCAAT.....GAGTGGCTACCTTCA	76351
369	erGluGluAspTyrMetAspLeuLeuTrpThrThrLeuSerGluPhePro	385
76352	TGACCTCCGACTTATGACACTGCTGTGGATCACCTGTCCGAGTTCTCT	76401
386	GlyValLeuValThrLeuTrpIleIleAspArgLeuGlyArgLysLysTh	402
76402	GGAATACTGTGACCATCAAGGTGGTCAAGCTGTTTGGCAAGAAAGAAC	76451
402	rMetAlaLeuCysPheValIlePheSerPheCysSerLeuLeuLeuPheI	419
76452	CATTGTTCTCAGTACCTGGCGCTGGTCTGTGCACTCTGGTCTGTAT..	76499
419	leCysValGlyArgAsnValLeu.....	426
76500GTGCGTGGAAAGGTAGTAGTGTGCACACTCAATTGATCATCTGA	76544
427ThrLeuLe	429
76545	TAATCAATCCAACTCCTCCTCAATCCTAGTCGCTTCTCCACTTCGGTAAC	76594
429	uLeuPheIleAlaArgAlaPheIleSerGlyGlyPheGlnAlaAlaTyrV	446
76595	CCTCTTCAATGACGCGCGACGATATCCGGCATCTTCCAGGCCATTTATG	76644
446	alTyrThrProGluValTyrProThrAlaThrArgAlaLeuGlyLeuGly	462
76645	TGTACATCCGGAATATACCCAGCTGCCTTGAGATCGTCGGAGTTTCG	76694
463	ThrCysSerGlyMetAlaArgValGlyAlaLeuIleThrProPheIleAl	479
76695	GGCTGTTCCGTGCTGGCAGACTCGGCGCATGCTGACCCCTTTCGTGGC	76744
479	aGlnValMetLeuGluSerSerValTyrLeuThrLeuAlaValTyrSerG	496
76745	GCAGTGTGTATGAGCTCGTCCAGAAATTCAGGCCATGTCACGTAACGAA	76794
496	lyCysCysLeuLeuAlaAlaLeuAlaLaserCysPheLeuProIleGluThr	512
76795	TCGTGGGCTGCTGGCTCCATAGCTCGCTCTTCTGCCAGGGAACACT	76844
513	LysGly	514
76845	GTGGC	76850
seq_name: gb_htg:AC020509		
seq_documentation_block:		
LOCUS	AC020509	157851 bp DNA HTG 03
DEFINITION	Drosophila melanogaster, *** SEQUENCING IN PROGRESS **	
ACCESSION	AC020509	pieces.
VERSION	AC020509.1	GI:6664388
KEYWORDS	HTG; HTGS_PHASE2.	
SOURCE	fruit fly	
ORGANISM	Drosophila melanogaster	
REFERENCE	1 (bases 1 to 157851)	
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Brachy	
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	

AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA

COMMENT This sequence was identified as CDH:10213488 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
1..157851
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

BASE COUNT 41166 a 56910 c 37402 g 42373 t
ORIGIN

alignment_scores:
Quality: 984.00 Length: 605
Ratio: 2.589 Gaps: 19
Percent Similarity: 62.810 Percent Identity: 39.835

alignment_block:
US-09-911-667A-2 x AC020509 ..

Align seg 1/1 to: AC020509 from: 1 ... 157851

```

41 ValHisValGly.LeuGluaIaValGIuL.iASP...AspGlyAlaIaV 56
|||||:||||| |||||:|||||:|||||:|||||
107085 ATCCACATGGAGCTGAAGACGGTGTATAAATCGGATGGTAGTGTT 107134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
56 alProlysGluPhe.....AlaAsnProThrAspAsp 66
|||||:|||||:|||||:|||||:|||||
107135 CGAAGAGAAATTTCACATTTTCGGTAACAAGTCACCTTTCACAGAT 107184
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
67 ThrPheMetValGIuAspAlaValGIuAlaIIeGlyPheGlyLysPheGI 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
107185 ACCTACACTGTTTCAGCAGCGATCAATGCCTTCGGTTTCGGATGTTCCA 107234
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
83 nTrpLysLeuSerValLeuThrGlyLeuAlaTrpMetAlaAspAlaMeGL 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
107235 CGTCAAGCTGTCTGCTCTGGTGGGGCTTCGTGGATGTCGACTCATCG 107284
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
100 LuMetMetIleLeuSerIleLeuAlaProGlnLeuHisCysGluTrpArg 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
107285 AAATGCCAATTCACACATATTGGGCCCGCTTTGTGTCGCGAGTGGAAAT 107334
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
117 LeuProSerTrpGlnValAlaLeuLeuThrSerValIaPheValGlyWe 133
::: :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
107335 GTCACCAAGTTCAGCAGCATCCGTCACAAACAGTGGTGTITGGGCAT 107384
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
133 MetSerSerThrLeuTrpGlyAsnTlleSerAspGlnTrpGlyArgL 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
107385 GATGTTGAGTTCAGTTTTGGACACAGCTGAGCAACAGATATGGCCGA 107434
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
150 ys.....ThrGly.LeuLys..... 154
|| |||||:|||||
107435 AGTCGGTAAGACTGACTGGTGAATCAATAAGCTAATAGCTACTAAAATGCC 107484
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
155 .....IleSerValLeuTrpThrLeuTyTrpG 164
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
107485 ACCATTTCAGGCCCTCACATGTTTCGGAGTACTACTAGTGCTG...TATA 107531
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
164 IyIleLeuSerAlaPheAlaProValTyTrpSerTrpIleLeuValLeuArg 180
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
107532 GCATTCTCAGCTCGGTGGCTCCAGTACGCCGTGGCTCTGACCCTAAGA 107581
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
181 GlyLeuValGlyPheGlyIleGlyGlyValProGlnSer..... 193
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
107582 GGFTCTGAGGCTTCGCATPCGGCTCGGTTCCACAGTCGTGGAGTAATTCC 107631
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
194 .....V 194

```

Tue Mar 12 15:54:23 2002

107632 ATGGAGTAGGAGGTACCTCTCCACCTAAACCTTTCTTTAAACCCAAAGAG 107681
194 aThrLeuTyrAlaGluPheLeuProMetLysAlaArgAlaLysCys... 209
107682 TTACCTTGACCTGAGTTCTGCGCCACCAAGCAGCAAGGAAAGTGGCTC 107731
209 209
107732 GTATTGATGGAGCTGAGTACGTGCCACCGTGTGCTAAACGATTATTGA 107781
210 11eLeuLeuLeuGluValPheTTPAlaIleClyThrValPheGlu 224
107782 ATATATTGACTGGTTTCTAGCTCTCTGGCCCTTCGGAGCTTGCTTTCAA 107831
225 ValValLeuAlaValPheValMetProSerLeuGlyTyrArgTrpLeuLeu 241
107832 GTGGTCTAGCACTGGTTGTGTATCCCTATTACGGATGGCGAGGTCTCTT 107881
241 u11eLeuSerAlaValProLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 255
107882 GGCCCTTTCGGCTACACCCCTACTCATTTTACCATCTCTATCGCCGGTTC 107931
255 255
107932 GTATTCTCTCGCTCTCTTCAGAAATATCTTAATCAAAATGTATCTTTAAAT 107981
256 Phe...TTPleuProGluSerAlaArgTyrAspValLeuSerGlyAsn 270
107982 CACTACAGTGGCTGAGTGAATCGCTCTGCTATCTTACTTCTTACAACGGTCTAC 108031
271 GlnGluLysAlaIleAlaThrLeuLysArgIle..... 281
108032 AATGATAAGGCCATCAAGGTCTCTGGAACAGGTGAGTGGTAACCTTTATGGA 108081
282AlaThrGluAsnCLYA 287
108082 GTGATTATACCTCTAAACCACTACGATTGTAGATAGCTCACACACACAG 108131
287 laPro.MetProLeuGlyLysLeuIleIleSerArgGlnGluAspArgGI 303
108132 CGCACATGCTGATGGTGGCTGATGGCGGAGGAGGAGCCAGTTCGCG 108181
303 yLys...MetArgAspLeuPheThrProHisPheArgTyrTrpThrLeuL 319
108182 GGAGAGCTTTATAGATCCCTGCTCAGCCGAGCCGCTGTACAGACCACTATCC 108231
319 euLeuTyrPheIleTrpPheSerAsnAlaPheSerTyrTyrGlyLeuVal 335
108232 TGTGTGCTTCTCTGCTGGCTGGCTGGCTTTTGTCTACTACGGCCGTGTG 108281
336 LeuLeuThrThrGluLeuPheGlnAlaGlyAspValCysGlyIleSerSe 352
108282 CTGGTCACGACCGAGCTGCTGGTGGCC..... 108308
352 tarGlyLysAlaValGluAlaLysCysSerLeuAlaCysGluTyrLeuS 369
108309 AGGAACAAGGAGAGCCATCCCAAT.....GAGTGGCTCACCTTCA 108348
369 erGluGluAspTyrMetAspLeuLeuTrpThrThrLeuSerGluPhePro 385
108349 TGACCTCCGACTTATGACCTGTGTGGATCACCTCTCGAGTTCCT 108398
386 GlyValLeuValThrLeuTrpIleLeuAspArgLeuGlyArgLysLysTh 402
108399 GGAATACTGCTCACCATCAAGTGGTCAAGCTGTTTGGCAAGAAGAAAC 108448
402 rMetAlaLeuCysPheValIlePheSerPheCysSerLeuLeuLeuPheI 419
108449 CATTGTTCTGACGACCTACCGCGCTGTTCTGTGACACTCTGTGCTGAT... 108496
419 leCysValGlyArgAsnValLeu..... 426
108496 426

108497GTGCGTGAAAGGTAGTAGTGGTGACACTCAATTGATCATCTGA 108541
427ThrLeuLeu 429
108542 TAATCAATCCAATCCCTCCTCCAAATCCTAGTGGCTTCTCCACTTCCCGTAAC 108591
429 uLeuPheIleAlaArgAlaPheIleSerGlyGlyPheGlnAlaIaTyrV 446
108592 CCTCTTCATTGCGACGGCGCATATCCGCATCTTCCAGGCCATTATG 108641
446 alTyrThrProGluValTyrProThrAlaThrArgAlaLeuGlyLeuGly 462
108642 TGTACACTCGGAAATATACCACTGCTTGAGATCCGTCGGAGTTTCG 108691
463 ThrCysSerGlyMetAlaArgValGlyAlaLeuIleThrProPheIleAl 479
108692 GGCTGTTCCGTGCTGGCGAGACTCGGCCCATCTGACCCCTTTCTGGG 108741
479 aGlnValMetLeuGluSerSerValTyrLeuThrLeuAlaValTyrSerG 496
108742 CGAGTGTCTGATGACTCGTCCAGAAATTCAGCCCATGTCACGTACGCA 108791
496 lYcysCysLeuAlaAlaLeuAlaSerCysPheLeuProIleCltThr 512
108792 TCGTGGCGCTGCTGGCTCCCATAGCTCGCTCTTCTCCCGAGGCAACT 108841
513 LysGly 514
108842 GTTGGC 108847
seq_name: gb_in:AE003462
seq_documentation: block:
LOCUS Al_11462 300542 bp DNA INV 04-OCT-2000
DEFINITION Drosophila melanogaster genomic scaffold 142000013386038 section 11
of 15, complete sequence.
ACCESSION AE003462 AE002575
VERSION AE003462.1 GI:7291637
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Drosophila
REFERENCE Eukaryota: Metazoa; Arthropoda: Tracheata; Hexapoda: Insecta;
AUTHORS Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 300542)
Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos,G.L., Abril,J.F., Aghayani,A., An,H.J.,
Andrews-Pfannkuch,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkov,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dunn,P., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunlop,P., Durbin,K.J., Evans,J., Fierman,C., Ferrera,S.,
Fleischmann,W., Fostier,C., Gabrielian,A.E., Garg,N.S.,
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Correll,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kuip,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.T.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Liu,Y., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
Morrison,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,


```

mRNA      31845..>33894
          /gene="CG5543"
          /product="Fly7560"
          /db_xref="FLYBASE:FBan0005543"
          <31845..>33894
          /gene="CG5543"
          /map="60A2-60A2"
          /db_xref="FLYBASE:FBan0005543"
          /db_xref="FLYBASE:FBgn0034908"
          31927..33894
          /gene="CG5543"
          /note="CG5543 gene product"
          /codon_start=1

alignment_scores:
  Quality: 984.00      Length: 605
  Ratio: 2.589        Gaps: 19
Percent Similarity: 62.810 Percent Identity: 39.835

alignment_block:
US-09-911-667A-2 x AE003462/rev ..

Align seg 1/1 to reverse of: AE003462 from: 1 to: 300542

41 ValHisValGly.LeuGluAlaValGluLeuAsp...AspGlyAlaAlav 56
   ::::::::::::::::::::
239236 ATCCACATTGGAGCTGACAGCGGTCTCTATAAATCCGGATGCTAAGTGGT 239187
   ::::::::::::::::::::

56 a1ProLysGluPhe.....AlaAsnProThrAspAsp 66
   ::::::::::::::::::::
239186 CGGAAGAGGAATTTTCACACATTTCCGGGTAAACAGTGCACCTTTCACAGAT 239137
   ::::::::::::::::::::

67 ThrPheMetValGluAlaValGluAlaValGlyPheGlyLysPheG1 83
   ::::::::::::::::::::
239136 ACCTACATGTTTCAGCAGCGGATCAATGCTTCGGTTTCGGATGTTTCA 239087
   ::::::::::::::::::::

83 nTrpLysLeuSerValLeuThrGlyLeuAlaTrpMetAlaAspAlaMetG 100
   ::::::::::::::::::::
239086 CGTCAAGCTGTGCTCTTGGTGGGCTCTGCTGGATGTCGACTCCATGG 239037
   ::::::::::::::::::::

100 LuMetMetLeuSerIleLeuAlaProGlnLeuHisCysGlnTrpArg 116
   ::::::::::::::::::::
239036 AAATGCCATTTTCACCATATTTGGGCCCTCTTGTCTCGCAGTGGAAAT 238987
   ::::::::::::::::::::

117 LeuProSerTrpGlnValAlaLeuLeuThrSerValValPheValGlyMe 133
   ::::::::::::::::::::
238986 GTCACCAAGTTCACAGCAGCATCCGCTCAACACAGTGGTGTGTTTGGGCAT 238937
   ::::::::::::::::::::

133 tMetSerSerThrLeuTrpGlyAsnIleSerAspGlnTrpGlyArgL 150
   ::::::::::::::::::::
238936 GATGTTGAGTTCAGCTTTTGGACACAGCTGAGCAACAGATATGCGCGGA 238887
   ::::::::::::::::::::

150 ys.....ThrGly.LeuLys..... 154
   ::::::::::::::::::::
238886 AGTCGGTAAGACTGACTGTAATCAATAGCTAAATAGCTACTAAATGCC 238837
   ::::::::::::::::::::

155 ..... 164
   ::::::::::::::::::::
238836 ACCATTTCAGGCCCTCACATGTTTCGGAGTACTACTAGTGGCTG...TATA 238790
   ::::::::::::::::::::

164 lYtleLeuSerAlaPheAlaProValTyrSerTrpIleLeuValLeuArg 180
   ::::::::::::::::::::
238789 GCATTCTACGTCGGTGGCTCCAGTTACGCCCTGGCTCTCGACCTTAAGA 238740
   ::::::::::::::::::::

181 GlyLeuValGlyPheGlyIleGlyGlyValProGlnSer..... 193
   ::::::::::::::::::::
238739 GGTCTGTAGGCTTCGCCATCGCTGCTGCTGCCAGTCCAGTCTGAGTAATCC 238690
   ::::::::::::::::::::

194 .....V 194
   ::::::::::::::::::::
238689 ATGGACTAGGAGTACCTCTCCACCTAAACATTTCTTTAACCCAAAAGAG 238640
   ::::::::::::::::::::

```

```

194 a1ThrLeuTyrAlaGluPheLeuProMetLysAlaArgAlaLysCys... 209
   ::::::::::::::::::::
238639 TTACCTTGTACGCTGAGTTTCTGCCCAACAGCACAAAGGAAAGTGGCTC 238590
   ::::::::::::::::::::
209 ..... 209
   ::::::::::::::::::::
238589 GTATTGATGGAGCTGAGTACGTGCCACCGTGTGCTAAACGCTATTATTGA 238540
   ::::::::::::::::::::
210 ....IleLeuLeuIleGluVal.PheTrpAlaIleGlyThrValPheGlu 224
   ::::::::::::::::::::
238539 ATATATTGACTGCTTTCAGTGTCTTGGGCTTGGGAGCTTGGCTTTCAA 238490
   ::::::::::::::::::::
225 ValValLeuAlaValPheValMetProSerLeuGlyTrpArgTrpLeuLe 241
   ::::::::::::::::::::
238489 GTGCTCTAGCACTGGTGTGTATCCCTATTACGGATGGCGAGTCTCTT 238440
   ::::::::::::::::::::
241 uIleLeuSerAlaValProLeuLeuPheAlaValLeuCys..... 255
   ::::::::::::::::::::
238439 GGCCTTTCGGCTACACCCCTACTCAITTTTACCACATCCTATCGCGGTTTC 238390
   ::::::::::::::::::::
255 ..... 255
   ::::::::::::::::::::
238389 GTATTCTCTGCTCTCTTTCAGAAATATCCTTTAATCAAAATGTTATTTAAAT 238340
   ::::::::::::::::::::
256 ....Phe...TrpLeuProGluSerAlaArgTyrAspValLeuSerGlyAsn 270
   ::::::::::::::::::::
238339 CACTACAGTGGCTCAGTGAATCGGCTCGCTATTACTCTTACAAACGGTCAC 238290
   ::::::::::::::::::::
271 GlnGluLysAlaIleAlaThrLeuLysArgIle..... 281
   ::::::::::::::::::::
238289 AATGATAAGGCCATCAAGGTTCTGGAACAGGTGAGTGTGAACATTTATGGA 238240
   ::::::::::::::::::::
282 .....AlaThrGluAsnGlyA 287
   ::::::::::::::::::::
238239 GTGATTATACGTTATAAAACCTACAGTTGTAGATAGCTCAACAACCAACAG 238190
   ::::::::::::::::::::
287 laPro.MetProLeuGlyLysLeuIleIleSerArgGlnGluAspArgG1 303
   ::::::::::::::::::::
238189 CCCCACATGCTGATGGTGGCTGATGGCCGACGACGAGCCAGTTGGCG 238140
   ::::::::::::::::::::
303 yLys...MetArgAspLeuPheThrProHisPheArgTrpThrLeuL 319
   ::::::::::::::::::::
238139 GGAGAGCTTTAGATCCCTGCTCAGCCCGAGCTGTACAGAGCCACCATCC 238090
   ::::::::::::::::::::
319 euLeuTrpPheIleTrpPheSerAsnAlaPheSerTyrTyrGlyLeuVal 335
   ::::::::::::::::::::
238089 TGTGTGTGTTCTCTGCTGGCTCGGCTTTTGCTACTACGGCTGGTG 238040
   ::::::::::::::::::::
336 LeuLeuThrThrGluLeuPheGlnAlaGlyAspValCysGlyIleSerSe 352
   ::::::::::::::::::::
238039 CTGGTCACGACCGAGCTGCTGCTGGCC..... 238013
   ::::::::::::::::::::
352 rArgLysLysAlaValGluAlaLysCysSerLeuAlaCysGluTyrLeus 369
   ::::::::::::::::::::
238012 .AGGAACAAGAGAGACCCATCCCAAT.....GAGTCGTCACCTTCA 237973
   ::::::::::::::::::::
369 exGluGluAspTyrMetAspLeuLeuTrpThrLeuSerGluPhePro 385
   ::::::::::::::::::::
237972 TCACCTCCGACTTTATGGACCTGCTGGATCACCCTGTCCGAGTTTCCT 237923
   ::::::::::::::::::::
386 GlyValLeuValThrLeuTrpIleIleAspArgLeuGlyArgLysLysTh 402
   ::::::::::::::::::::
237922 GGAATACTGCTGACCATCAAGGTGTCAAGCTGTTTGGCAAGAGAAAAC 237873
   ::::::::::::::::::::
402 rMetAlaLeuCysPheValIlePheSerPheCysSerLeuLeuPheI 419
   ::::::::::::::::::::
237872 CATGTGTTCTGACGATACCTGGCGCTGGTGTCTGTGCACCTCTGGTGTCTGAT... 237825
   ::::::::::::::::::::
419 leCysValGlyArgAsnValLeu..... 426
   ::::::::::::::::::::
237824 .....GTCGGTGGAAAGGTAGTAGTGGTGCACACTCAATTGATCATCTGA 237780
   ::::::::::::::::::::
427 .....ThrLeuLe 429
   ::::::::::::::::::::

```

```
237779 TAATCAATCAATCTCTCTCCAACTAGTCGCTTCTCCACTTCCGTAAC 237730
429 uleuPheIleAlaArgAlaPheIleSerGlyPheGlnAlaAlaIleVal 446
237729 CCTCTTCATTCGACGGCAGCATATCCGGCATCTTCCAGGCCATTTATG 237680
446 alTyThrProGluValTyThrProThrAlaThrArgAlaLeuGlyLeuGly 462
237679 TGTACACTCCGGAATATACACAGCTGCTTGCAGATCCGTCGGAGTTTCG 237630
463 ThrCysSerGlyMetAlaArgValGlyAlaLeuIleThrProPheIleAl 479
237629 GGCTGTTCGCTGCTGCCAGATCGGCCCATGCTGACCCCTTTCGTGGC 237580
479 aGlnValMetLeuGluSerSerValTyThrLeuThrLeuAlaValTySer 496
237579 GCAGGTGCTGATGGACTGCTCAGAGATTCAGGCCATGTCCAGGTACGCA 237530
496 lyCysCysLeuLeuAlaAlaLeuAlaSerCysPheIleProIleGluThr 512
237529 TCGTGGGCTGCTGGCTCCATAGCTGCTTCTGCCAGGGAAACT 237480
513 LysGly 514
237479 GTTGGC 237474
seq_name: gb_pr:HSM802697
seq_documentation_block:
LOCUS HSM802697 2104 bp mRNA PRI '8-JUN-2000
DEFINITION Homo sapiens mRNA; cDNA DKFZp761H039 (from clone LKfZp761H039).
ACCESSION AL359592
VERSION AL359592.1 GI:8655658
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 2104)
Blum,H., Bauersachs,S., Mewes,H.W., Weill,B. and Wiemann,S.
Direct Submission
Submitted (15-JUN-2000) MIPS, Am Klopferspitz 18a, D-82152
Martinsried, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by LMU (Ludwig Maximilians University,
Munich/Germany) within the cDNA sequencing consortium of the German
Genome Project.
This clone (DKFZp761H039) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://www.mips.biochem.mpg.de/proj/cDNA/.
Location/Qualifiers
1..2104
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp761H039"
/clone_lib="761 (synonym: hamy2). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
/tissue_type="amygdala"
1..437
/gene="DKFZp761H039"
<1..437
/gene="DKFZp761H039"
note="similarity to SV2 related protein (Rattus
norvegicus)"
/codon_start=3
/product="hypothetical protein"
/protein_id="CAB94878.1"
/db_xref="GI:8655659"
FEATURES
source
gene
CDS
/translation="LCFVIFSFCLLIFCVGRNVLTLLFIARAFISGGQAAVYVT
PEVYPTATRALGLGTCMARVAGALITPFAQVMESSVYLTLLAVYSGCCLLAALASC
FLPIETKRGRLQESSHREMGQEMVGRMGHGVTRSNSSGQE"
polyA_signal
polyA_site 2090
BASE COUNT 462 a 557 c 579 g 506 t
ORIGIN
alignment_scores:
Quality: 734.00 Length: 144
Ratio: 5.133 Gaps: 0
Percent Similarity: 99.306 Percent Identity: 99.306
alignment_block:
US-09-911-667A-2 x HSM802697
Align seg 1/1 to: HSM802697 from: 1 to: 2104
405 LeuCysPheValIlePheSerPheCysSerLeuLeuPheIleCysVa 421
3 CTGTGCTTTGTTCATCTCTCTTCGAGGCTCTGCTGTTTATCTGT 52
421 lGlyArgAsnValLeuThrLeuLeuPheIleAlaArgAlaPheIleS 438
53 TGAAGAATAATGTCTCACTCTGTTACTTCTTCATTGCAAGAGCGTTATT 102
438 erGlyGlyPheGlnAlaAlaIleValTyThrProGluValTyThr 454
103 CTGGAGGCTTTCAGCGGCATATGTTTACACACCTGAGGTCTACCCACG 152
455 AlaThrArgAlaLeuGlyLeuGlyThrCysSerGlyMetAlaArgValG 471
153 GCAACCGGGGCTCGGCTGGCCACCTGCGCCAGGATGCGGCAAGAGTGG 202
471 yAlaLeuIleThrProPheIleAlaGlnValMetLeuGluSerValT 488
203 TGCTCTCATCTCTGCTTTCATCGCCAGGCTGATGCTGGAATCTCTGT 252
488 yLeuThrLeuAlaValTySerGlyCysCysLeuLeuAlaAlaLeuAla 504
253 ACCTGACTCTGGCAGTTTACAGTGGCTGCTGCTGCTGCTGCTGCTGCC 302
505 SerCysPheLeuProIleGluThrLysGlyGlyLeuGlnGluSerSe 521
303 TCTGCTTTTCCCATTCAGACCAAGGCCGAGGACTGCAGAGTCCAG 352
521 rHisArgGluThrGlyGlnGluMetValGlyArgGlyMetHisGlyAlaG 538
353 CCACCGGGAGTGGGGCCAGGAGATGTCGGCCGAGGAATGCACGCTGCAG 402
538 lyValThrArgSerAsnSerGlySerGlnGlu 548
403 GTGTTACCAAGTCGAACTCTGCTCTCAGGAA 434
seq_name: gb_pl:AB026645
seq_documentation_block:
LOCUS AB026645 81020 bp DNA PLN 07-MAY-1999
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, pl clone: MGH6,
complete sequence.
ACCESSION AB026645
VERSION AB026645.1 GI:4757401
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui PI
clone:MGH6.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (sites)
REFERENCE Nakamura,K.
AUTHORS
TITLE Structural Analysis of Arabidopsis thaliana Chromosome 3. II
```

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (28-APR-1999) to the DDBJ/EMBL/GenBank databases.
Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of
Gene Structure 2; Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail: ynakamu@kazusa.or.jp, Tel: 81-438-52-3935(ex. 7443),
Fax: 81-438-52-3934)

```

FEATURES             Location/Qualifiers
     source            1..81020
                        /organism="Arabidopsis thal
                        /strain="Columbia"
                        /db_xref="taxon:3702"
                        /chromosome="3"
                        /clone="MGH6"
BASE COUNT           24700 a 15544 c 15238 g 25538 t
ORIGIN

```

```
alignment_scores:
  Quality: 723.50
  Ratio: 2.055
  Percent Similarity: 52.381
  Length: 672
  Gaps: 16
  Percent Identity: 28.423

alignment_block:
  US-09-911-667A-2 x AB026645
  ..
  Align seq 1/1 to: AB026645
  from: 1 to: 81020
```

[illegible]

67716 TGTGTCAATGGTCAATCAATCAATGCTAAATGGACTTTGTTTGGGTCAT 67766
151 GlyLeuLysIleSerValLeuTrpThrLeuT 151
67766 ATTCTTCTGTTTATCATGCATTTGGCTTCATTTGATTTTGATGTTTGTG 67815
152 GlyLeuLysIleSerValLeuTrpThrLeuT 162
67816 CCTGAATTCCTCAGGAAGGGTTTATCATAACTACGCGTGGTGACTTTTG 67865
162 yTrpGlyIleLeuSerAlaPheAlaProValTrpSerTrpIleLeuVal 178
67866 TAGCTGGTTCCTTGAGCCATTTTCACCTAACTACATGCTGTTGATCAT 67915
179 LeuArgGlyLeuValGlyPheGlyIleGlyValProGlnSerValTh 195
67916 CTTTCGCTGTTGGTTGGTCTGGGTGGAGAGCTCCTGTTCTGTCTTC 67965
195 rLeuTyrAlaGluPheLeuProMetLysAlaArgAlaLysCysIleLeuL 212
67966 TTGGTATCTTGAATTCATCTGCACCAAGCGGAGGACTTGGATGGTTG 68015
212 euIleGluValPheTrpAlaIleGlyThrValPheGluValValLeuAla 228
68016 TTTTCTCAGCTTCTCGGACCGTTGGAACCAATTTTGGAGCTTCCCTTGC 68065
229 Val...Phe..... 230
68066 TGGGTATAACTCATGCCCTTCTTCTGTTTATT, ICAGTATTAATTTGGACG 68115
231 ValMetPr 233
68116 CTAGCTTTTCACATCACCATGCAATGCAATCTTTTGCAGTTAGTCATGCC 68165
233 oSerLeuGlyTrpArgTrpLeuLeuIleLeuSerAlaValProLeuLeuL 250
68166 AAGATTGGGTTTGGAGTGGTTACTTTCGATCTCTCTCGGTAGCTTCTCGT 68215
250 euPheAlaValIleCysPheTrpLeuProGluSerAlaArgTrpAspVal 266
68216 TACTTCTCTGTTCTTATAGTGGACATCTGAGCTCTCCGCTACTTAAT 68265
267 LeuSerGlyAsnGlnLysAlaIleAlaThrLeuLysArgIleAlaTh 283
68266 CTACAAGCGGAAAGCCGAAGCTCTGCCATATTAGAGAAGATTGCGAG 68315
283 rGluAsnGlyAlaProMetProLeuGly..... 292
68316 GATGAACAAACCCCAACTACTCCGGGTGTTCTTCAGTTTCAGAGTTGGAGA 68365
292 292
68366 CCGAGCTGGAAGAGATAAAACATTTCCACAGAAAAATCTCATCTTCTC 68415
293 LysLeuIleIleSe 297
68416 AAGCTGGAGAAAGTGGAGAGCGGTTGCTGTGCTAAGATTGTACTAAA 68465
297 rArgGlnGluAspArgGly...LysMetArgAspLeuPheThrProHisP 313
68466 ACCTGATAAGNACACAGGTTCTCATTTACTAGCTTCTTCTATCTCAACGC 68515
313 heArgTrpThrThrLeuLeuLeuTrpPheIleTrpPheSerAsnAlaPhe 329
68516 TAATGAACGGGACTTTGTCTACTATGGGTGTGTCTTTTGGAAATGCGTTT 68565
330 SerTyrTrpGlyLeuValIleLeuThrGluLeuPheGlnAlaGlyAs 346
68566 GCTTACTATGGCGTTGCTCTCTCACCACCGAGCTGCAACAAATCCATAA 68615
346 pValCysGlyIleSerSerArgLysLysAlaValGluAlaLys..... 360
68616 CCGCTGTCTA...TCCAACCGGAAACAGTTGCGAAATTTCAACAGCATGTT 68661

```
361 .....CysSerLeu..... 363
68662 AACACAGACGCTTTCATTGCCAGTTTCGCAGGTACATCAATATATT 68711
364 .....AlaCysGluTyrLeuSerGluGluAs 372
68712 CTCTTATCAAGAAATCCAAAGCAGTTGTACTTCTTCAAAATAAGAAA 68761
372 pTyrMetAspLeuLeuTrp.....ThrThrLeuSerGluPheProGly 387
68762 AAGTCATAATCTTATTTTGTGATCTGTGGCGCTTCAGAGTTTCCAGGTC 68811
387 aLeuValThrLeuTrpIleIleAspArgLeuGlyArgLysLysMet 403
68812 TACTGATATCAGCAGCAATGGTGATAGACTTGGACGAAAGCTCAATG 68861
404 AlaLeuCysPheValIlePheSerPheCysSerLeuLeuPheIleIcy 420
68862 GCA.....TCAATGTGTTCACATGTTGCATTTTCTCTCTCCCTCTGTT 68905
420 sValGlyArgAsn.....ValLeuThrLeuLeuLeuPheIleAlaArg 435
68906 ATCCATCAGTCTCCATTCATACACCGCTTCTCTTCGGCGCGCGTA 68955
435 laphelleSerGlyGlyPheGlnAlaAlaTyrValTyrThrProGluVal 451
68956 TCTGCATCTCGGAGCTTTTCACAGTGTGTTCATATATATGCTC 69005
452 TyrProThrAlaThrArgAlaLeuLeuGlyThrCysSerGlyMetAl 468
69006 TATCCACGCGCAGTGAGACACCGAGTAGGAGTAGGAGCTCGTAGG 69055
468 aArgValGlyAlaLeuThrProPheIleAlaGlnValMetLeuGluS 485
69056 AAGAAATTGGAGGAATCTTATGTCACACTCGTAGCC...GTTGGATTAGTTC 69102
485 erSerValTyrLeuThrLeuAlaVal.....TyrSerGlyCysCysLeu 499
69103 ACGGATCCATCAGACAGTCGCTGCTCTCTGTTGAGGTGTGATCTT 69152
500 LeuAlaAlaLeuAlaSerCysPheLeuProIleGluThrLysGlyGly 516
69153 GTGTGAGGAATTTGCGTTTCTCTCTTCCCTTTCGAGACAGTGGTCTGTA 69202
516 yLeuGlnGluSer 520
69203 CCTGACAGATAGT 69215
seq_name: em_pl:AC024128
seq_documentation_block:
ID AC024128 standard; DNA; PLN; 81020 BP.
XX AC AC024128;
XX AC AC024128.1
XX 02-MAR-2000 (Rel. 62, Created)
XX 02-MAR-2000 (Rel. 62, Last updated, Version 1)
XX Arabidopsis thaliana chromosome III p1 MGH6 genomic sequence, complete
DE sequence.
XX HTG.
XX Arabidopsis thaliana (thale cress)
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
XX eurosids II; Brassicales; Brassicaceae; Arabidopsi.
XX [1]
RN 1-81020
RP
```

```
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B., Wu D.,
RA Matti R., Rønning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III p1 MGH6 genomic sequence";
XX Unpublished.
XX [2]
RN 1-81020
RP Lin X.;
RT Submitted (24-FEB-2000) to the EMBL/GenBank/DBJ databases.
RL The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD
RL 20850, USA
XX SPTREMBL; Q9LD39; Q9LD39.
DR SPTREMBL; Q9LDB0; Q9LDB0.
DR SPTREMBL; Q9LDD1; Q9LDD1.
DR SPTREMBL; Q9LDF8; Q9LDF8.
DR SPTREMBL; Q9LDK4; Q9LDK4.
DR SPTREMBL; Q9LDM2; Q9LDM2.
DR SPTREMBL; Q9LDS0; Q9LDS0.
DR SPTREMBL; Q9LDY9; Q9LDY9.
DR SPTREMBL; Q9LDZ8; Q9LDZ8.
DR SPTREMBL; Q9LE42; Q9LE42.
DR SPTREMBL; Q9LE44; Q9LE44.
DR SPTREMBL; Q9M7W1; Q9M7W1.
DR SPTREMBL; Q9M7W2; Q9M7W2.
DR SPTREMBL; Q9M7W3; Q9M7W3.
DR SPTREMBL; Q9M7W4; Q9M7W4.
DR SPTREMBL; Q9M7W5; Q9M7W5.
DR SPTREMBL; Q9M7W6; Q9M7W6.
XX Address all correspondence to:atetigr.org
CC P1 clone MGH6 is from Arabidopsis chromosome III and is near the
CC molecular marker .
CC The orientation of the sequence is from SP6 to T7 end of the P1
CC clone.
CC Genes were identified by a combination of three methods: Gene
CC prediction programs including GRATL (available by anonymous ftp
CC from arthur.epm.ornl.gov), Genefinder (Phil Green, University of
CC Washington), Genscan (Chris Burge,
CC http://gnomic.stanford.edu/~chris/GENSCANW.html), and NetPlantGene
CC (http://www.cbs.dtu.dk/netgene/cbsnetgene.html), searches of the
CC complete sequence against a peptide database and the Arabidopsis
CC EST database at TIGR (http://www.tigr.org/tdb/at/at.html).
CC Annotated genes are named to indicate the level of evidence for
CC their annotation. Genes with similarity to other proteins are named
CC after the database hits. Genes without significant peptide
CC similarity but with EST similarity are named as 'unknown' proteins.
CC Genes without protein or EST similarity, that are predicted by more
CC than two gene prediction programs over most of their length are
CC annotated as 'hypothetical' proteins. Genes encoding tRNAs are
CC predicted by tRNAscan-SE (Sean Eddy,
CC http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
CC identified by RepeatMasker (Arian Smit,
CC http://ftp.genome.washington.edu/RM/RepeatMasker.html).
XX Key Location/Qualifiers
FH 1. 81020
FT source
FT /chromosome="III"
FT /db_xref="taxon:3702"
FT /organism="Arabidopsis thaliana"
FT /cultivar="Columbia"
FT /clone="MGH6"
FT mRNA
FT join(<426..518,645..>1034)
FT /gene="MGH6.1"
FT CDS
FT join(426..518,645..1034)
FT /codon_start=1
FT /db_xref="SPTREMBL:Q9M7W6"
FT /note="unknown protein"
FT /gene="MGH6.1"
FT /protein_id="AAF35940.1"
FT FT
```

```

FT FT /translation="MMHEFRPLTANDTIPGGPHTSNPTPTSLHAEVWTLRIKRTVYS
FT FT SKKYPDRRELANGRKQOOSNYEAVINFGDNESSTNNMNVNRECKGNVRSVFUL
FT FT QQTYPQHQNPILMTTTHVDSQHFHSNDNIHETVETWPDELRSVVEFAFPSELS"
FT FT complement(2496..2802)
FT FT /note="ATTIR1673A1ATTIR1673A ATTIR1673A , a non-autonomous
FT FT DNA transposon - a consensus.."
FT FT complement(3046..3122)
FT FT /rpt_family="(CAA)n"
FT FT join(complement(13663..>13709),complement(13169..13565),
FT FT complement(<12513..13076))
FT FT /gene="MGH6.3"
FT FT join(complement(13663..13709),complement(13169..13565),
FT FT complement(12513..13076))
FT FT /codon_start=1
FT FT /db_xref="SPTREMBL:O9LDD1"
FT FT /note="hypothetical protein"
FT FT /gene="MGH6.3"
FT FT /protein_id="AAF35941.1"
FT FT /translation="MAVEAHILNPLFSSNREMIHPVEASGVVYNTOMRYGTVTPTNPTV
FT FT ECOTSLFNPINISPDRLVHQSMKPTIQSDVSSLTFNSDNNVDFLRPVSSRRKREES
FT FT VILNPSAYWQIQKNPTDPLMFLGODLSSNVQOHFDIDRLISNHVRMMMEIEKKKTO
FT FT GRRIEAVQGLMKTLRAKDELINHLGKLNLFLEEKVKSILCVENOIWRDVAQSNKATVN
FT FT ALKNSIQVLAVERNRWEEPTTVADDAQSCGSDGDSBEERKMLAGEAODTKMCR
FT FT VGMSCRCGKEASVLLPLCRHMLCSVCGSSLNFCPICKSPKTLASLHVNLSS"
FT FT complement(13950..14024)
FT FT /rpt_family="(TAAA)n"
FT FT join(<18851..19090,19255..19381,19476..19614,19697..19785,
FT FT 19881..>20002)
FT FT /gene="MGH6.4"
FT FT join(18851..19090,19255..19381,19476..19614,19697..19785,
FT FT 19881..20002)
FT FT /codon_start=1
FT FT /db_xref="SPTREMBL:O9LDY9"
FT FT /note="hypothetical protein"
FT FT /gene="MGH6.4"
FT FT /protein_id="AAF35942.1"
FT FT /translation="MASSTGLTVAGALLAGDFRLPAVSSLIPIKHTSSSLNSLRDLSS
FT FT PYNKWRSLSGKILTSLNSRKFVAKGEADGFLSNVSDTDHFDLRLKYGKVFERS
FT FT TYKCSPTAEVDDDAELAFAVELAKASVDYKAGDIKLVFKVPLVYVWTRFFIATAFSFR
FT FT QIDAIGSRMDRLAEKKYGVANGDVKPNSTLLDFDGVVTHLFLPQRTFYNLDFYGN
FT FT AMOIELPFEDOSQPRN"
FT FT join(complement(21254..22367),complement(<20664..21091))
FT FT /gene="MGH6.5"
FT FT join(complement(21254..22298),complement(20664..21091))
FT FT /codon_start=1
FT FT /db_xref="SPTREMBL:O9M7W5"
FT FT /note="unknown protein"
FT FT /gene="MGH6.5"
FT FT /protein_id="AAF35943.1"
FT FT /translation="MMAGNQMPTLGRVKVCDLVPSECLPSDSYKLAVTTLQSIAQYSA
FT FT AIIOFPASDGLRSGLDSARLYFHQRDSTPATNNMHTNDSQENKTSQYVADPQSQW
FT FT ESEYRPLPTPEPSNMEFPFAGLPDIFALLGKAARVLDALGFYLNLRSCPFTEILD
FT FT NVPLRNCVSSVSVCCYARSPFHGAQHSLTEDEOLILYSDBDHLDSLISQVKSQD
FT FT KAGLHIDHGMWILYDVDPQVAVYVGLQATYAGVYSPAVHRTDLNSLQSGIEG
FT FT RESLAPKMPKMTNLSVDSCEMRAAGHVEAQFQLPVSDDFMQRSHNSDLNFRQTLQS
FT FT FTVPOSDGSMKOLKRRKSDRCKPLPPSKRLLEAORVLKERVQEIADKKGILRFC
FT FT NLKECNSHNSMPCANIRRETCWPHGVPEVPHDLPNKAKIETLEPGWSETHDM
FT FT EFLSUTEAQGNQHVTCNHSPPNTPFPFAET"
FT FT join(complement(24815..>25204),complement(24182..24731),
FT FT complement(24256..24496),complement(23677..24177),
FT FT complement(<23184..23578))
FT FT /gene="MGH6.6"
FT FT join(complement(24815..25204),complement(24582..24731),
FT FT complement(24256..24496),complement(23677..24177),
FT FT complement(23184..23578))
FT FT /codon_start=1
FT FT /db_xref="SPTREMBL:O9LDF8"
FT FT /note="unknown protein"
FT FT /gene="MGH6.6"
FT FT /protein_id="AAF35944.1"
FT FT /translation="MPQYSGTGOHCETAAYSFSWPTSSRLSNAAEERANVFSNLOKEE
FT FT DDDVEPSPVSTPKQGRATTLLELMTIRAFHSMKLRCLYSLGTAIGFRIIRRGVLTDP
FT FT AIIVFVSRKVKHMKOWLSPLQCLPTALEGAGGIMCDVVEFSYFEGEPDHPQTPKQTFITD

```

```

FT FT IVDHLQSDPFIQSGSOVASOFTCTLCATVRSOTGQRQGVFTNHRVAVNLDPYSOKM
FT FT FHPPLPAGPVYLCABERATSFITDDLWFGIFACTNPETFVRADCAFIIPRADDYDLNR
FT FT VTSYGVGVGEIVKATLQELQSPGVSKQVQKAVRSGSLTGTGLVLAALFYNDERGV
FT FT CSLTDLFVVGENHRSFPDLQEGDSLIVMKGEKARPIGIIMGSTGSGRKLKLVGECPC
FT FT EWTGTVDIGLRLTHLOLDLITTDGLKAAVOQRAASTTGMSSHVADUSSPPYVYKKE
FT FT KRSPKEKLEASGLPQVOHIDLEERIEIKGGAPSVHEOFMPTFGCGCSASAMPETARED
FT FT LVAGFTNCSGCDLVCGLRQDDGAKRRRTQVTNERMRAE"
FT FT complement(26470..26521)
FT FT /rpt_family="(GA)n"
FT FT <26932..>27192
FT FT /gene="MGH6.7"
FT FT 26932..27192
FT FT /codon_start=1
FT FT /db_xref="SPTREMBL:O9LE44"
FT FT /note="hypothetical protein"
FT FT /gene="MGH6.7"
FT FT /protein_id="AAF35945.1"
FT FT /translation="MAKNKDDIKYATAOKLSEDEAIRVSKHGTPLEGKIAESEPVE
FT FT LFSSAQRLEKCKEQAASGDOTQIORDIKDKGTRTDDSPR"
FT FT complement(<31715..>32860)
FT FT /gene="MGH6.8"
FT FT complement(31715..32860)
FT FT /codon_start=1
FT FT /db_xref="SPTREMBL:O9LD39"
FT FT /note="unknown protein"
FT FT /gene="MGH6.8"
FT FT /protein_id="AAF35946.1"
FT FT /translation="MASGCVKNVGTSPSHSTSSKMSLTRESQPLAPALENEDPVDDFE
FT FT FLEDPVTMLSADELFSGLKPLKFSVGTVPPEEKPTTSVHTAVKPCRRLMEISGVV
FT FT DYLFSPRAPRCTVWRRELLGLKRLAKTOQEASASSSSRLSSSSPNPKTASFRHLNRS
FT FT SKSTAQOPSHPPGKSDSDILESSSTISSRLSSSSSGHEDDLPRLSLDLKNKPG
FT FT TPNPFAERSFAHHHLRNONOPKPRRHTQVDESTESSIESKVMVTADSPRLNASGKI
FT FT VFHGLERSSSPONTGGPRMKLIHGMPSHSANVRITPVLNVVPSLSRSGPKSGLFFG

```

alignment_scores:

```

Quality: 723.50 Length: 672
Ratio: 2.055 Gaps: 16
Percent Similarity: 52.381 Percent Identity: 28.423

alignment_block:
US-09-911-667A-2 x AC024128 ..

Align seg 1/1 to: AC024128 from: 1 to: 81020

49 GluLeuAspAspGlyAlaAlaValProLysGluPheAlaAsnProThrAs 65
|||||
67249 GAAATGGCAGATGGG.....AA 67265

65 pAspThrPheMetValGluAspAlaValGluAlaIleGlyPheGlyLysP 82
||| |||:|||||: |||:|||||:|||||
67266 CACAAGATTCACGGTGGATGAAGCTCTTGTGCGATGGGATTTGGCAAGT 67315

82 heGlnTrpLysLeuSerValLeuThrGlyLeuAlaTrpMetAlaAspAla 98
||||| :||| :|||:|||||:|||||:|||||
67316 TCCAAATTTACGCTCTCTTACGTGGAAATGGGTGGGTTCAGACGCC 67365

99 MetGluMetMetIleLeuSerIleLeuAlaProGlnLeuHisCysGluTr 115
|||||:|||||:|||||:|||||:|||||
67366 ATGGAGATGATGCTCTCTCTTCTTCTGACCTGCTGCTCAATCTCTCTG 67415

115 pArgLeuProSerTrpGlnValAlaLeuLeuThrSerValValPheValG 132
|||||:|||||:|||||:|||||:|||||
67416 GAATCTCTCTGACCCAGGAAGCTTAATAACTAGCGTCGTTTTCGTCG 67465

132 LysMetSerSerThrLeuTrpGlyAsnIleSerAspGlnTrpGly 148
|||||:|||||:|||||:|||||:|||||
67466 GCATGCTCATTTGGTCTACTCTTGGGGCATGTTCTTCGATAACATGCG 67515

149 ArgLys.Thr..... 151
|||||: |||
67516 CGCAGTAGCTTTTATCATACTTGTGTTGTTCTTGCATATAAGTTTCTTGTA 67565

151 ..... 151

```


67566 CCTCAACTAGAAAGGGGACGGCTGAAATGTTCCGTAAATGTTTCATATTCT 67616
151 151
67616 TATGAGCTCTTGAATTGAATGATGATCTCGACCTTGTCTAATGCTCAATTC 67665
151 151
67666 AGACTGTTTAGCTTATGACCATCCTTGTGTTTCATCCGATTTCTGTGTGCT 67715
151 151
67716 TGTGTCATAGTTCATCAATCAATGCTAAATGGACTTTGTTTGGGTCAT 67765
151 151
67766 ATTCTTCTGTTTATGATGCATTTGGCTTTCATTTTATGTTTTCATGTTTGTG 67815
152 GlyLeuLysIleSerValLeuTrpThrLeuT 162
67816 CTTTGAATTCCTCAGGAAGGGTTTATCATTAACCTCGCGTGGACTTTTG 67865
162 yTyrglyIleLeuSerAlaPheAlaProValTyrserrTrpIleLeuVal 178
178 178
67866 TAGCTGCTTCTTGAGGCGCATTTTACACCTAATCATGTTGTTGATCAAT 67915
179 LeuArgGlyLeuValGlyPheGlyIleGlyValProGlnSerValIth 195
67916 CTTTCGCTGTTGGTGTGTTGGGTGGAGGAGGCTCCTGTTCTTCGCTTC 67965
195 rLeuTyrrAlaGluPheLeuPrometLysAlaArgAlaLysCysIleLeuL 212
67966 TTTGGTATCTTGAATTCATCATCTGCACCAAGCGAGGACTTTGGATGGTTG 68015
212 eulleGluValPheTrpAlaIleGlyThrValPheGluValValLeuAla 228
68016 TTTTCTCAGCTTCTGGACCGTGTGAACCATATTTTGAGGCTCCCTGCT 68065
229 Val..Phe..... 230
68066 TGGGTATAACTCATCGCTTCTTCTTCTTATTTTCCAGTATAATTTGGAG 68115
231 ValMetPr 233
68116 CTAGCTTTTCATCACCATCGCAATCGAATCTTTTGCAGTTAGTCATGCC 68165
233 oSerLeuGlyTyrrArgTrpLeuLeuIleLeuSerAlaValProLeuLeuL 250
68166 AAGATTGGGTTGCAGGTGCTTACTTGCATCTCTTCGGTACCTTCTTCGT 68215
250 euPheAlaValLeuLysPheTrpLeuProGlnSerAlaArgTyrrAspVal 266
68216 TACTTCTTGTGTTATAGGTGCACATCTGAGTCTCCTCGGTACTTAAT 68266
267 LeuSerGlyAsnGlnGluLysAlaIleAlaThrLeuLysArgIleAlaIth 283
68266 CTACAAGCGCGGAAGCCGAAGCTCTTGCCATATTAGAGAAGATTTCGAG 68315
283 rGluAsnGlyAlaProMetProLeuGly..... 292
68316 GATGAACAAACCAACTACCTCCGGGTGTTCACGTTTCAGATTCGAGAG 68366
292 292
68366 CCGAGCTGGAAGAGATAAAACATTTCCACAGAAAAATCATCATCTTCTC 68415
293 LysLeuIleIleLeu 297
68416 AAGGCTGGAGAAAGTGGAGAGCGGTGTGCTGAAGATTGTACTAAA 68466
297 rArgGlnGluAspArgGly...LysMetArgAspLeuPheThrProHisP 313
313 313

seq_documentation_block:	198677 bp	DNA	BCT	22-NOV-1999
LOCUS	AE001863	Deinococcus radiodurans R1 section 2 of 2 of the complete		
DEFINITION	Chromosome 2.			
ACCESSION	AE001863			
VERSION	AE001863.1			
KEYWORDS	GI:5460670			
SOURCE	Deinococcus radiodurans.			
ORGANISM	Deinococcus radiodurans			

REFERENCE AUTHORS	Bacteria: Thermus/Deinococcus group; Deinococcales; Deinococcus. 1 (bases 1 to 198677) White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D., Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L., Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C., Makarova, K.S., Aravind, L., Daly, M.J., Fraser, C.M. et al. Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1 Science 286 (5444), 1571-1577 (1999) 20036896
TITLE	1. 198677
JOURNAL MEDLINE REFERENCE AUTHORS	2 (bases 1 to 198677) White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D., Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L., Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C., Makarova, K.S., Aravind, L., Daly, M.J., Minton, K.W., Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S., Smith, H.O., Venter, J.C. and Fraser, C.M. Direct Submission Submitted (02-NOV-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA Location/Qualifiers 1. 198677 /organism="Deinococcus radiodurans" /strain="R1" /db_xref="taxon:1299" /chromosome="2" complement(91..1926) /gene="DRA0206" complement(91..1926) /gene="DRA0206" /note="similar to GB:AE000520 percent identity: 43.92; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="oligodeoxytase F, putative" /protein_id="AAFL2423.1" /db_xref="GI:6460718" /translation="MLARACPLRGACPRSCRPAPSTCPCMERERWARLPMTTQ SNLEAVEHKLMVPADDAOWAYAPFAALQEAADLTAGDVPAWLERNNLSARLOGSGS KFLSTHADLHTDDDAIOORYQTFVAELVPOAERASQALTEKLLAVPGVPGPDALNIR RFQDAALFPEANVDIGVTINRHSVITGNQVLOGQOELTLPQAKOOMPSPDRA QREAAARLASNAPELDAVMDLIRTRROLAHNADANEDFRYRWKLEDVDTYP FOCLDFHRSVDEWVPLLSGLVGDIAEROMRESGLDLESRPDKMTHAYCNYPTTNEPVL GADLEALAOOTYVESLQPDLAARFQRMRESGLDLESRPDKMTHAYCNYPTTNEPVL MNVGTAEDVRVLFHEVGHAFHGFYSGDROPLVNRWSPTEFVEIPSMAMEFLTLDLH SHVFSDDDELARRKQLEGVIAFLPWAQMDAFQHLIXAEAPEDVTIADLDAKWLILD RTFHPFDMSGLDESVRAGKGMHYTHIFQVFPYTYIYACMLAAVAIWRARQDPAAAL GRYKDALRLGNTVSPLELYRAAGAEFRFDEYIRGLMGFLKEQLG" complement(2005..3645) /gene="DRA0207" complement(2005..3645) /gene="DRA0207" /note="similar to PID:1653945 percent identity: 49.28; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAFL2422.1" /db_xref="GI:6460717" /translation="MNEFVLSPLSSVAHALNRAKSLKEEMSKSRSLALLGALLL SACSPENGQAPAPQPLNTEARPGTSGKVRGLWYDAFGPGIKTPEEYSTLVDTAR AMNNTLYVQVGRDCCYCNNAAMPRTDDEPVPAGFPDLADYIAKAKADIEVNAWII TTALNWSVTPKSAEHAFTNHPGNAAGDNWLTVKADGTVKAGADWVLDPGHPDAAAY IRNNYSVAKNYAVDGFIMFDVRYPDYNPVGQYQWGYNETALARYAEATGTGTPDP ADPQWSARROOVTNLVRETAALAVKAVPGVATAAATITYEGPADAEAFAAASPYAA VQDMLTWKSGVLDONVMNMYKRDVPLQAGFDWNTFAAGLRQYPTHEQVSGAA IYLDNDASTVSOIKTKQAAGLWAGYSYRTPDKDVNEGKGAEEVLADLKLRLTDAGG PFPTEVHAPLPAASVRLSGELPAARPGHSHVELVRAGGTWNVVAATSDGQGRYG FLNLPGQADVVRVDKILLPNVQATPGRVTHLEPITLN" complement(3751..4860) /gene="DRA0208"
gene	complement(3751..4860) /gene="DRA0208" /note="similar to PID:1001216 PID:1001269 percent identity: 66.97; identified by sequence similarity; putative" /transl_table=11 /product="peptide ABC transporter, permease protein" /protein_id="AAFL2421.1" /db_xref="GI:6460716" /translation="MTGGETRTPLQMAWRRYKSRPGVIAGWVLIALYVIALSPFL APYNTTQHEGYEYORPOTVHLIHQRLHRPFFVVGSTTKRDPVTFAKTTPVDTARPLP VQFPVKGPPSYFELGLOSRRHLFGVPDGDGDEYGEKVPVFELFGDDOIGLDFSRTLVG ACQSVTVGVGVLLISFALGILLGVSGVFGWVDTLQORVVEVLLSFPRLPILIALAT LIPARMPSTWYTGIVAVLALIGWASLARVQGVLAARNLDYSAQAQMGASNLRLVI LRHTPLNSSFLLATLALPGYILGESTLSLFGIGTKEPMTSMGLLRDASKLEWVS SYPMILWPLGVFISVLAFFNFMGDALRDAADQSR" complement(4857..5843) /gene="DRA0209" complement(4857..5843) /gene="DRA0209" /note="similar to PID:1001492 PID:1001545 percent identity: 62.93; identified by sequence similarity; putative" /transl_table=11 /product="ABC transporter, permease protein" /protein_id="AAFL2420.1" /db_xref="GI:6460715" /translation="MLCCHMLNVTLRRLGLMPTLLVLSVLCFLVKLPQSGFIDQYL EDPFSKETVANITRGLDQPAWKQYLNMTGVVTRLDGYSFASGAPVTSIIGERL GWTFIAALTALTLIAVPLGIYAFNRYGPAQITANFLGVYGLAIPDFIAALLLVV LVRLQGGTVNGGLFRPDMIDAPWSPAKVLDLHLAIPVLAVGLEGVAGLMQRMRAST LDVLSQDXTIRAKGLPGQVRIWKHVRNAINPLISLAGLSLPTLSISITVMS LPTIGLLYDLSLLNKDQYTMATLLMSLAFLLLLIGNLLSLDALAWADPRVRS" complement(5964..7804) /gene="DRA0210" /note="This region contains an authentic frame shift and is not the result of a sequencing artifact; similar to GB:AE000657 percent identity: 51.19; identified by sequence similarity; putative" 7764..8603 /gene="DRA0211" 7764..8603 /gene="DRA0211" /note="similar to GB:AL009126 percent identity: 60.78; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="transcriptional regulator, GntK family" /protein_id="AAFL2419.1" /db_xref="GI:6460714" /translation="MWSNNKTFSELVRKMYGCAVSSPAPLPPEPPAWVVALDPAAP AYVQLAQLRRRIESELGASAGTAPAEALAAHGVSRVTLRQGLGLLEAGLGLRRR GSGTFVAGTPHTSRQLCLLSFSDVPAQGVFARVLSFTQGRPTPQALSLGLTAA EQYVRLRLRTSDGPLAVEDSTLPAALIGPLTAQDVTDASLYGLLQRRHLPAPAIR HLRAQNAEALAHLLGVPLCAALLTTERVSWLASGRPVEYARACRYGDRYDFVMDLRG EDH" 8603..9806 /gene="DRA0212" /note="This region contains an authentic frame shift and is not the result of a sequencing artifact; similar to GB:LA2023 SP:p44861 PID:1005712 PID:1220844 PID:1205001 percent identity: 56.04; identified by sequence similarity; putative" 9857..10774 /gene="DRA0213" 9857..10774 /gene="DRA0213" /note="similar to GB:U00096 SP:p76535 PID:1788768 percent identity: 63.89; identified by sequence similarity; putative" /codon_start=1
gene	8603..9806 /gene="DRA0212" /note="This region contains an authentic frame shift and is not the result of a sequencing artifact; similar to GB:LA2023 SP:p44861 PID:1005712 PID:1220844 PID:1205001 percent identity: 56.04; identified by sequence similarity; putative" 9857..10774 /gene="DRA0213" 9857..10774 /gene="DRA0213" /note="similar to GB:U00096 SP:p76535 PID:1788768 percent identity: 63.89; identified by sequence similarity; putative" /codon_start=1
gene	8603..9806 /gene="DRA0212" /note="This region contains an authentic frame shift and is not the result of a sequencing artifact; similar to GB:LA2023 SP:p44861 PID:1005712 PID:1220844 PID:1205001 percent identity: 56.04; identified by sequence similarity; putative" 9857..10774 /gene="DRA0213" 9857..10774 /gene="DRA0213" /note="similar to GB:U00096 SP:p76535 PID:1788768 percent identity: 63.89; identified by sequence similarity; putative" /codon_start=1
gene	8603..9806 /gene="DRA0212" /note="This region contains an authentic frame shift and is not the result of a sequencing artifact; similar to GB:LA2023 SP:p44861 PID:1005712 PID:1220844 PID:1205001 percent identity: 56.04; identified by sequence similarity; putative" 9857..10774 /gene="DRA0213" 9857..10774 /gene="DRA0213" /note="similar to GB:U00096 SP:p76535 PID:1788768 percent identity: 63.89; identified by sequence similarity; putative" /codon_start=1
gene	8603..9806 /gene="DRA0212" /note="This region contains an authentic frame shift and is not the result of a sequencing artifact; similar to GB:LA2023 SP:p44861 PID:1005712 PID:1220844 PID:1205001 percent identity: 56.04; identified by sequence similarity; putative" 9857..10774 /gene="DRA0213" 9857..10774 /gene="DRA0213" /note="similar to GB:U00096 SP:p76535 PID:1788768 percent identity: 63.89; identified by sequence similarity; putative" /codon_start=1

```

/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF12418.1"
/db_xref="GI:6460713"
/translation="MTDPRRTQVHPDYADLTLPADALIAAALADUUGLAVRAVQAA
POLTAALNAEVPOLGERGRVYVAGAGTSGRLGVLDATDELTPFSPPERAVPLIAGE
RAIRONVEAGEDAERGDPOAVNIGPDVLIIVAASGTTPYVLGAARSGRAAGALT
VGLANNPGLPLAAVDCPVLDTGPEILSGTRLKAGTAOKIALNTLSSALMVLKJL
YGNLMYDVRATNAKLEDRRLRVLQHTAGTADADAQAALSECAGSVKTLVWLKIGLGA
QEAQRLEGAGGAROVIGEAGEALTSAS"
/complement(10796..11398)
/genes="DRA0214"
CDS
complement(10796..11398)
/genes="DRA0214"
/notes="similar to GB:M99166 SP:P30849 PID:148264 GB:U00096
PID:1651499 percent identity: 54.36; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="trp repressor binding protein Wrba, putative"
/protein_id="AAF12417.1"
/db_xref="GI:6460712"
/translation="MTAPVKVLAIVYSSGTGYAMAQEAAGRAAGAEVRLKVRRT
APQDIDGODAMKANLEAMKDPPEATPADLEWAEIYVSSPTREGGATSQMRAEIDL
GLWSSGKLANKTFESAMTSAQNVNGQETLIQTLXTAMHWGAVLTTPGYTDEVIKFS
GNNPYGASVTANGQPLENDRASIRHVRROVELTAKLLG"
/11664..12059
/genes="DRA0215"
CDS
11664..12059
/genes="DRA0215"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF12499.1"
/db_xref="GI:6460794"
/translation="MGSGLHPROLGHLDRLSDELPMFGLNPREVAFAFLKAGLINAAG
LTGMWGTPHPALGDPAAPFALLSDNEQARLDGLVRLVVAHHLLCTLGLAPLIDPND
TERVWQRIETRPPLFVDEVMAGREDEGA"
/12068..13069
/genes="DRA0216"
CDS
12068..13069
/genes="DRA0216"
/notes="similar to GB:U00096 PID:1788770 PID:1799860
PID:1799864 percent identity: 54.32; identified by
sequence similarity; putative"

alignment_scores:
Quality: 644.50 Length: 464
Ratio: 2.170 Gaps: 5
Percent Similarity: 64.009 Percent Identity: 30.603

alignment_block:
US-09-911-667A-2 x AE001863 ..

Align seg 1/1 to: AE001863 from: 1 to: 19877

70 ValGluAspAlaValGluAlaIleGlyPheGlyLysPheGlnTrpLysLeu 86
77115 GTGACCGCGTGTGGACGACCTGTGGCGCTGGCGCGCTTCAGTGGAAAGCT 77164
86 userValLeuThrGlyLeuAlaTrpMetAlaAspAlaMetGluMetMetI 103
77165 GCTGCCCATCTGTGGCTGACCTGTGGCGCGCGGACGATGGAGGTGCTCC 77214
103 LeuSerIleLeuAlaProGlnLeuHisCysGlnTrpArgLeuProSer 119
77215 TCATGGGCTTCGCGCTCCCGGCATCAGTGCAGGCTTCAGTTCGCCAAA 77264
120 TrpGlnValAla.....LeuLeuThrSerValPheValGlyMetMe 134
77265 GCCTGCCCGCGCGGACGATGCTCTCACCAGCGACATTTTCGGGGGATGCT 77314

```


[illegible]

seq_name: qb_htq:H0702G05

seq documentation block:

LOCUS	H0702G05	134092 bp	DNA
seq_documentation_block:			

DEFINITION *Oryza sativa* chromosome 4 c

IN PROGRESS **, in ordered pieces.

ACCESSION AL442105

VERSION
KEYWORDS
SOURCE

AL442105.1 GI:10241609
HTG: HTGS PHASE2.

Orvza sativa.

SOURCE ORGANISM

Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eubartoideae; Oryzaceae; Oryza.

REFERENCE

[illegible]

TOURNAY.

2 (bases 1 to 134092)
Hong, G. F.

AUTHORS
TITLE:

Submitted (15-SEP-2000) Hong G.F., National Centre for Gene Research, Chinese Academy of Sciences, 500 Caobao Road, Shanghai, 200233, CHINA

COMMENT

This is not a completed record, but the order is confirmed. Caps' length are less than 500bp. The first gap is between 15752bp and 15753bp, the second gap is between 35844bp and 35845bp, the third gap is between 71662bp and 71663bp. This record will be updated as soon as the sequence is finished.

- * NOTE: this is a working draft sequence.
- * This sequence will be replaced
- * by the finished sequence as soon as it is available and
- * the accession number will be preserved.

DECLASSIFIED

```
1. .134092
   /organism="Oryza sativa"
   /sub_species="indica"
   /variety="Guangluai4"
   /db_xref="taxon:4530"
   /chromosome="4"
   /clone_lib="BAC-HindIII"
   /clone="H070205"
   /accession="cg10"
```

RECEIVED

3/2/74 B 29132 C 29130 Y 30513 C

ORIGIN

scores:			
Quality:	572.00	Length:	897
Ratio:	1.658	Gaps:	20
Similarity:	38.462	percent identity:	21.182

Percent 3

block:

-667A-2 x H0702G05 ..

1/1 to: H0702G05 from: 1 to: 134092

hrPhmetValGluAspAlaValGluAlaIleGlyPheGlyLysPheGI 83
::: |||::|||::: :: |||::|:::|
CGATACCGCCCATACCCCCCTTCATCAGCTTTTGGGAGGTACCA 68502

TripLysLeuservAlleuthrGlyLeuAlaTrpMetAlaAspAlametG 100
:: ::|||::::: :||||:|||||
CCTATAGCGTGGTAGACGCCGCCTTGTCATCAAGTTTCCGAGGTAAACC

ATATTGATCTTTTCATATGCCGGGTCGCTTGATTGCAGAAAGCAATGG 68555

euProSerTrpGlnValAlaLeuLeuThrSerValValPheValGlyme 133
GATGATGCTACTATCATTTGTTGGTCCATCAGTACAGTTAGAATGGAAG 68602

TTTACTTCATCAGGAAGCATGATTACAAGCATGTTTGTGTTGGAAT 68652
MetSerSerSerThrLeuTrpGlyAsnIleSerAspGlnTyrGlyArg. 149

11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533

68653 GCTAATAGGAGCTTACACTTGGGTGTAGTTTCAGACAACATACGAAGGA 68702
149 149
68703 GGCAGGTCCATCGTTTACTCCAATCACTGATTTTGTTCATGACACAT 68752
149 149
68753 TATGCAATAAATTTGATCTCTTGGTGATGCTTATGCTTTAGCTTAGCA 68802
149 149
68803 ATGGGTACAAAAAGATGGTGTGAACCTTGTTCATGACACAGCTTGTTC 68852
149 149
68853 CAGTTGATAGTTTATAGCCATAGTTATTAAAGACCGGACCGGCGGCTC 68902
149 149
68903 GGACCGGAAAAACAGAACTGGAGCCTCTGCCGGTTCCGTCCAGTTAAA 68952
149 149
68953 AGACCGGATATGCAATCAAGCCGGTATAACCCGGTTAGGCTGTCGGGTTT 69002
150 LysThrGlyLeuLysIle..... 155
69003 TCATGAAAAACCGCTATAGAACCGGAGAAAAAACCATGGTTGAACCGGCC 69052
155 155
69053 TTGCCAGGAGTGGGTGCATCAAGGGAGCAGTCGGGGGAAAAATGAGGGA 69102
155 155
69103 AAAAGAGCCCTCACCACCCAGGCTATCATCTGTGTGTTTGAAG 69152
155 155
69153 GAATGAGTAATATATTTAATAACTATTGAACCGGTTAGACAAATGGTTC 69202
155 155
69203 AACCGGTTGAACCACTGAACCATCGAACCAAGAGCCTCACCGGTTCAATT 69252
155 155
69253 ACCGGTCCAGTCTCAATAACTATGTTTTAGCTAGTAACAACTGGC 69302
155 155
69303 AAAGGTCTTAGATAATTGATCAGTTAGACATTTCTATGTTACTTTCAGAA 69352
156 Ser.ValLeuTyr 159
69353 ATATCCTTGAACCACTCCCTCAAAATTTATGTTCTCGAAATCATATATTTT 69402
159 rp..... ThrLeuTyrTyr..... 163
69403 GGTGATGCTTATTTTCATGATATTTGCAACACACTACCTTGATATATTCCTCT 69452
163 163
69453 TGGTGTGCAGCACAGAGCTCATACATTCTGTTCATTTGTAATTTACAG 69502
164 GlyIleLeuS 167
69503 GAGAGGATTTCTTTTACCGGCATCGTGACTAGTGGAGCTGGATTCCTCA 69552
167 exAlaPheAlaProValTyrSerTrpIleLeuValLeuAArgGlyLeuVal 183
|||||..... |||||
69553 GTGCTTTTGCACCTAACTATGTGTCACACTAATTTCTCTAAGATTTCTAGTT 69602

184 GlyPheGlyIleGlyGlyValProGlnSerValThrLeuTyrAlaGluPh 200
|||||..... ||| |||||
69603 GGTATTGGATTGGAGGGACCTGCTTGGCATCTTGGTTCTTGGAAAT 69652
200 eLeuProMetLysAlaArgAlaLysCysIleLeuLeuIleGluValPhe 217
|||||..... |||||..... |||||
69653 TGTTCCTGCTCCAACTAGAGGAAGCTGGAGTGGTAGTATTTTCTGCCTTT 69702
217 rpAlaIleGlyThrValPheGlu..... 224
69703 GGACTGTTGGACCATCTTCGAGGCTTCTCTGCTGGGTATGATATTTT 69752
224 224
69753 TCCTGCTATTTTTAAACTGGAATATCTGACTTTTACTTTTGGTGGGTAT 69802
224 224
69803 CCAAGGTCAATAGAAGTGCCCTGTTCTGCTCATTTCTCATTCATGAAGC 69852
225 Val 225
69853 CTACATTCTCATTTGAAAAATGTATGTAACATATCTTTAGGCCCTGTCAAT 69902
226 ValLeuAlaValPhe..... ValMetProSerLe 235
69903 ATTGTAACATATCTTTTCATCATGCTAATTGACAGACAGTAAATGCCCAAT 69952
235 uGlyTrpArgTrpLeuLeuIleLeuSerAlaValPro..... LeuL 249
69953 TGGCTGAGGTGGTTGCTAGCATTTCTGCTGCCATCTTTCTGCTCTGCG 70002
249 euLeuPheAlaValLeuCysPheTrpLeuProGluSerAlaAArgTyrAsp 265
70003 TTTTATTTTATGTTATT..... ACACCAGAGTCAACCAAGGTTCTCTA 70043
266 ValLeuSerGlyAsnGlnGluLysAlaIleAlaThrLeuLysArgIleAl 282
70044 TGCATGAAGGCAGAGTAACAGAAGCCATGGATGTTACTGGAGAAAATGGC 70093
282 aThrGluAsnGlyAlaProMetProLeuGlyLysLeuIleIleSerArg. 298
70094 TAGATTGAACAATGTACAGCTCCCTCTGGAAGGCTTGATCTGACAAA 70143
298 298
70144 ACATTGAGCTAGATGAACCTTTCACATCTGAGTCTACACACTTCTTGCT 70193
299 GlnGluAspArgGlyLysMetLargAsp..... 307
70194 GATCCTGAAGAAGATGACACCATCATTCAGACAAAGTTTCCAGTTTGG 70243
308 LeuPheThrProHisPheArgTrpThrL 318
70244 AGGTTTGTCTCGCAGAGCTGCTGCGCCAAAATTTGATACGTGCAACTC 70293
318 euLeuLeuTrpPheIleTrpPheSerAsnAlaPheSerTyrTyrGlyLeu 334
70294 TGCCTCTGGATGACTTCTTTGGAAATGCTATTTCTATATTGGAAT 70343
335 ValLeuLeuThrThrGluPheGlnAlaGlyAspValCysGlyIleSe 351
70344 GTTCTGCTCACATCGGAGTTAAGTAACGGTAATAGGATATGTGA..... 70388
351 rSerArgLysLysAlaValGluAlaLysCysSer. LeuAlaCysGluTyr 367
70389 AAAGAGGAGGTGAATCACTACATCAATGATGCAAGCTTAT 70431
368 LeuSerGluGluAspTyrMetAspLeu..... LeuTrpThrThrLe 381
70432 ACAAAAACGTTTTTATATATCTAGCTTTGCGAGGTTTCTTTCTGTACATT 70481

381	u.....	381
70482	ATTAATTTCTCCAGTCGATATGATTCCTCCATTTTAAACAACGAATGCTGC	70531
382SerGluPheProGlyValLeuValThrLeuTrpIle	393
70532	TCTGTTTCTTGAACAGAGATTCAGGATCATTTTGTACCCCATGATT	70581
394	IleAspArgLeuGlyArgIysLysThrMetAlaLeuCysPheValIlePhe	410
70582	GTGATAGGATTGGTAGAAGCTGCAATGGCA.....TCAATGCTCTT	70625
410	eSerPheCysSerLeuLeuPheIleCysValGlyArg.....AsnValIle	426
70626	CACAAGCTGCTCTTATTCCTCCACTAATTTCTCCCGAAGGATATAC	70675
426	euThr....LeuLeuLeuPheIleAlaArgAlaPheIleSerGlyGlyPhe	441
70676	TAACAAGAGTCTCTCTATTGGAGCCCGGCTTTGCATATCTGCTAGCTTT	70725
442	GlnAlaIatYrValYrThrProGluVal.....	451
70726	ACAATTGTATACATATATGCTCCAGAGGTTGGTTCAGTCTGCTGCCITCA	70775
451	451
70776	CATGTTTCAGTAGTTCAGTCTCTACTAGATAGCACACATAATTTGTTTCATT	70825
451	451
70826	TCATCATGCTGACACAAATGGAAGATCAGATGTTAATGCTAACTGTTT	70875
452TyrProThrAlaThrArgAlaLe	459
70876	TCTTCTCTGCTTCAATTTTCAGATTATCCCAACCTCTGTCAGACAAC	70925
459	uGlyLeuGlyThrCysSerGlyMetAlaArgValGlyAlaLeuIleThrP	476
70926	CGGAATTGGCTGGGAGCTCTGTAGTAGGATTGGAGGATCCTCTGTC	70975
476	roPheIleAlaGlnValMetLeuGluSerSerValYrLeuThr....Leu	491
70976	CTCTGCTGCTGTTGCTCTAGTACACAGCTGCCAGCAGACAGCAGCATC	71025
492	AlaValYrSerGlyCysCysLeuLeuAlaAlaLeuAlaSerCysPheLe	508
71026	CTTCTCTTGGAGCTTGTAATCTCTCTCCGGTGTGCTGCTCATTTT	71075
508	uProIleGluThrIysGlyGlyLeuGlnGluSer	520
71076	CCCTTTTGACAAAAGGCTGTAGACTTAACGACACC	7112
seq_name: gb_ba:AE008141		
seq_documentation_block:		
LOCUS	AE008141	11286 bp DNA BCT 14-AUG-2001
DEFINITION	Agrobacterium tumefaciens strain C58 circular chromosome, section 199 of 254 of the complete sequence.	
ACCESSION	AE008141	AE007869
VERSION	AE008141.1	GI:15157408
KEYWORDS	Agrobacterium tumefaciens.	
SOURCE	Agrobacterium tumefaciens.	
ORGANISM	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae; Rhizobium.	
REFERENCE	1 (bases 1 to 11286)	
AUTHORS	Hinkle,G., Slater,S.C. and Goodner,B.	
TITLE	Complete Genome Sequence of Agrobacterium tumefaciens C58 (Rhizobium radiobacter C58), the Causative Agent of Crown Gall Disease in Plants	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 11286)	
AUTHORS	Hinkle,G., Slater,S.C. and Goodner,B.	

TITLE	Direct Submission
JOURNAL	Submitted (14-AUG-2001) Cereon Genomics, 45 Sidney Street, Cambridge, MA 02139, USA
FEATURES	Location/Qualifiers
source	1. .11286
	/organism="Agrobacterium tumefaciens"
	/strain="C58"
	/db_xref="taxon:358"
gene	118..1128
	/gene="AGR_C_4098"
CDS	118..1128
	/gene="AGR_C_4098"
	/note="(AP001509) BH0731
	unknown conserved protein"
	/codon_start=1
	/transl_table=11
	/product="AGR_C_4098p"
	/protein_id="AAK87995.1"
	/db_xref="GI:15157409"
	/translation="MGVMTGETLDEKAIVASAEKALSDIAAIRTEVSKVIFGOEKVV
	QNTLLAILSGHALLVGPGLAKTKLVITGLVDANRIQFTPDLPMSDILGSEVM
	DQDENGRRSFRFTKGPITFGOLLMADEINRASPTQSALQAQMOEYHITWAGOTVELPK
	PFHVLTQNPLEQEGTYPLPEAOLDRFLQVDVDPYELAAERQIILLDTTGTASGEARP
	VIGARLMEIQAIIROMPSDKYVDATILSLVRSARPCGHGKLTGKNVWMPGPGPRAGOS
	LMLTARARALYEGRLAPSLDDVHALAEPVLEHKNALTFARAEGMSVROVIAALVEQA
	KN"
gene	1159..2079
	/gene="AGR_C_4100"
CDS	1159..2079
	/gene="AGR_C_4100"
	/note="hypothetical protein"
	/codon_start=1
	/transl_table=11
	/product="AGR_C_4100p"
	/protein_id="AAK87996.1"
	/db_xref="GI:15157410"
	/translation="MPVASIGQIVDKTPGSEVLARAPQRAALVPDCMVEAKRIANTVT
	AGHGRKRKGIGENFQRPYAGESLSRDWRSARDHTYVDRDREAAHTIWLVA
	DMSPMMFKSTLGSVSKESALVLMALAEILABSGERICGPGIMEVSRNAAERLA
	AATMHAPLTSGMPTDSIMIGASDILVIGDFLDDAKNYMERISPLARGLRHVVIEAD
	PAELTFYSGRTEFSDPETGEKLVSGRAETITDYTRAYLARREALSSLRRLGNFV
	FRTHDLASEALVAVHMYLSGAPAGGGGHR"
gene	2076..4886
	/gene="AGR_C_4102"
CDS	2076..4886
	/gene="AGR_C_4102"
	/note="hypothetical protein"
	/codon_start=1
	/transl_table=11
	/product="AGR_C_4102p"
	/protein_id="AAK87997.1"
	/db_xref="GI:15157411"
	/translation="MSALPFVFTSPYILFGLLALPAITWLLRLTPRPKARVFPPLKI
	LATVLKREETPSKSPWLLTLRLALAAVIFALADPVVNPNNISAGSGLVILVVDNS
	WASAPDERRVATAQALIGDAERDRPVISFTADAEDHADVSTTAVALDKLAAKPK
	PLVPDRARTAEITAEALNGTAPGTLAYIADGVQTAQDESAMRTLASUSPAEFRIVRGD
	GKAVADITATNADNASTVTLRLDTAQAASTVNAQDSQGRILANSTPAPGAAGET
	TATVEAPELNRDFARLSIEGVSTAGAVHLLDDGFRRRVALLAGETGNDPQLIQLP
	YYISRALQPFVDLIPORADLAQALPEILQSNPSVIMADIGRLPQETTYVPLQRLAG
	GCTLIIRFAGPRLAAPADDPLVPVTLPGGERALGALSWAEPQPLADFPAGFAGMP
	KADGILVRQVLAIEPTDPAERTWASLADGTPLVLTNRNVEAGRIVLPHVSAESWSDL
	PISGHVRDMLRRIVOLSKAGGASANAPALPPFRLLTAEGAMSIEIGTARPLEMR
	AGQPRTGFDNPPGLYGEDGFVALNLLPAGATLRPLDTLSAAGVSTTNEALIGETAKS
	LRPALFIAAFNMLIADSLIVLFMNGAFARLPKARSATVLTLLALGAFAAASPTDARA
	DDPKGDSQIFERLDTTHAVRTGEDDVDIRISQGLQGESEFTLWRTTLDPQSPVGL
	DISKDELSPFLIYVPVSATAMPSSAISRIDAYMAGGTVLFDTRDQSSLDIGAT
	SANGERLQAILANDIDPLEVPEDHVLTRSFYLLTNFPCRYNGSLPWVESRGAAKT
	TLGSSSGDGVTPIMITGNDGAWAVDAQAPVLPVTPDEMREHAFRSGVNIMY
	MLTGNKYKQDVHPALLERLGO"
gene	4888..6972
	/gene="AGR_C_4105"
CDS	4888..6972
	/gene="AGR_C_4105"

```

/feature "note"hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_4105p"
/protein_id="AAK87998.1"
/db_xref="GI:15157412"
/translation="MTLNPAPLPWIVIALAVAALASFGLWGVGKAWLRLGLAIV
ALLAAAPLLTQEREPLSTVPVIVDRSQSDQVDRPMTDTALETLIKPLSRFPR
IEPRIVEVDNGESDPSQLESALSSAVSDSPSRVGAIFLSDQIHDINPLPNA
EQALGFAPVHGILITGKADEFORRIEIVRAPRFGIVNEEQULTLRFDGRRPAGGSA
EYTVKNGEIAATLOATPOQPTGFSFKVPVGGNNWMEFSVAELPGEVTAANRAVHLI
EGTRNNRLVLYSGEPHACERAWRLKSDASVDLVHFTILRPPEKQDQCTPINELSLI
APPTRELVEKINDPDLIIIDRYQHRGVLPILYDYIAQYVEKGALLIAAGPEHAGO
DSTMTPLASVLPQAQPTGQVHOAAEYPRLSEAGKKHPVTRGLESGNEPQWRWERT
ISVDPQGGTVMGDDGDPDLLVNRKQGRVAKMLSDQNLWARGFEFGGPHVALYKK
IAHMLKEPELEEALATARANGRTLOITRQITGDDPGQASVRFPFGKTNEMAFITTEP
GLYIRERRDEETGLFEIKNGDETTLVGVADPAEFKAMISTETLKPLADATKGSVH
RVADNSGRJALPDILPVRGDIRVAEDRMLIRLTDTETLVKGVNTLPLFAGFAGLAIL
LAFSALWREGR"
7017..7475
/feature "note"AGR_C_4106"
7017..7475
/feature "note"AGR_C_4106"
/feature "note"(L25828) ORF 12"
/codon_start=1
/transl_table=11
/product="AGR_C_4106p"
/protein_id="AAK87999.1"
/db_xref="GI:15157413"
/translation="MRPKSGAVRRLEKNTMNFVLSVDADAEEKAIRDPLVAYNIARF
GESDKRDLNITIRNDONSVTGVLGHITARGMLYVOLLFPVPEAMKGQCIAPKLLMAEE
EARKCGMAYIDTMPDALRTYERYGFTKIGSLGLSSQGSITWLEKRF"
complement(7485..8798)
/feature "note"AGR_C_4107"
complement(7485..8798)
/feature "note"AGR_C_4107"
/feature "note"probable sugar transporter - Deinococcus
radiodurans (strain R1)"
/codon_start=1
/transl_table=11
/product="AGR_C_4107p"
/protein_id="AAK88000.1"
/db_xref="GI:15157414"
/translation="MTAITNDALEREAGAGAYQRRIMGIFGLVWAADAMJVLAVGFTA
ASTAATFLVPAQALOTGTAFFGMLLGAAGFGLADRYGRRRLVITVACDALFQVL
SIFSPDTILLVLRFTLGAAGVGTLPVDYAMAEFLPAKNRMLVLELEGFWAVGTLLI
VALANGASLAGVADAWRYIAFVATPAVLGLGLRELVPESPLYLLRTGRSEFAKAV
NRMIVNGASLPDAGTGLFPQDVTGKGFISPAKLRQBSMMILAIWFLVSVSYKCVFTW
MPKLAGDGFGFVGYGFLVLAQAIPGYIAAYGVENKGRKPTLIGCLLSALGCL
LFTLASGAMIAASLLIMSALLIGTQWALYAFTEPYPTESRATGMAAGAMARLQGL
LAPSLIGYAIAGQFGVAGTIGFAGLLVLAATMINAETRQVALT"
complement(8888..9493)
/feature "note"AGR_C_4108"
complement(8888..9493)
/feature "note"AGR_C_4108"
/feature "note"probable acetyltransferase VC0655 {imported} -
Vibrio cholerae (group O1 strain N16961)"
/codon_start=1
/transl_table=11
/product="AGR_C_4108p"
/protein_id="AAK88001.1"
/db_xref="GI:15157415"
/translation="MSPHPSGSGFLGFKSKHDLVLYLTEDASHDAIIDLINEEAFGPGRH
TRAAARIEQGHDRSLUSFVCADDEGTIASVRMTPVLAVAGKGMHMLGFLAVRPSKNNH
GIGRELVRITAIAAARRKSSAAILIGDPPYIMPLGFPEKFAVAAALDFPGVPDPNRVLVY
PVGDDVHQRLKLGKIGWRKCEAAMPAAAAAEDESEEPRLVYG"
9586..10746
/feature "note"AGR_C_4109"
9586..10746
/feature "note"AGR_C_4109"
/feature "note"HYPOTHETICAL 37.4 KD PROTEIN IN EXUR-TDCC
INTERGENIC REGION (O328)"
/codon_start=1

```



```
166 LeuSerAlaPheAlaProValTyrSerTrpIleLeuValLeuArgGlyLeu 182
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8495 CTTTCGATATTTACCGCATTTACCATTTCCGCTCGCTCGCTGCTTCT 8446

182 uValGlyPheGlyLeuGly...ValProGlnSerValThrLeuTyrA 198
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8445 CACCGAGCTCGGTGGGCGGCATTTCCCGTGAGTATGCGATGATGG 8396

198 IaGluPheLeuProMetLysAlaArgAlaLysCysIleLeuLeuIleGlu 214
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8395 CGAATTCCTGCCTCCCAAAATCGCGGCGCTGCTGCTGTTCTTCGAA 8346

215 ValPheTrpAlaIleGlyThrValPheGluValValLeuAlaValPheVa 231
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8345 GGTTCCTCGGCGGTGGCACACTGATC...GTGGCACTCGCGCTTGGG 8299

231 ImetProSerLeuGly...TrpArgTrpLeuLeuIleLeuLeu 244
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8298 TGCCAGCTCTCGGCGGTGGCTGATGCTCGGCTATATTTTTCGGTGA 8249

244 eAlaValProLeuLeuLeuPheAlaValLeuCysPheTrpLeuProGlu 260
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8248 CGGCTTCCTCGGCGGTCTTCTGCTCGGCTTCTGCTGCTGCTGCGAA 8199

261 SerAlaArgTyrAspValLeuSerGlyAsnGlnGluLysAlaIleAlaTh 277
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8198 TGCCGCTCTACCTGCTCGGACGGGCGATCCGAAGAAGCAAGCCGT 8149

277 rLeuLysArgIleAlaThrGluAsnGly...AlaProMetProLeuGlyL 293
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8148 CGTCAACCGCATGCTGGTGGTGAACGGCGCTGCACCGCTCGATCGCGAA 8099

293 ysLeuIleIleSerArgGlnGluAspArgGlyLysMetArgAspLeuPhe 309
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8098 CGGCGCTTTTC...CAGCTCAGCTGACCAAGCGCGCGGATATTT 8055

310 ThrProHisPheArgTrpThrThrLeuLeuLeuTrpPheIleTrpPheSe 326
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8054 TCGCGCGCTCGCGACGCGATGATGCTCGCATCTGCTGCTTCT 8005

326 rAsnAlaPheSerTyrTyrGlyLeu...ValLeuLeuThrThrGluLeuP 342
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8004 CGTCTCCCTTCTTATATGCGCTTCCACCTGATGCTCGCGCAACTG. 7956

342 heGlnAlaGlyAspValCysGlyIleSerSerArgLysLysAlaValGlu 358
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7955 .....GCAGGGACGCTTTCGGTTTC..... 7935

359 AlaLysCysSerLeuAlaCysGluTyrLeuSerGluAspTyrMetAs 375
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7934 .....GTGCGCGTTACGGGTT 7918

375 pleuLeuTrpThrThrLeuSerGluPheProGlyValLeuValThrLeuT 392
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7917 TCTGTGTGTGGCTGGCGGCGATCCCGGTTATGCAATGCTGCT 7869

392 rPileIleAspArgLeuGlyArgLysLysThrMet...AlaLeuCysPhe 407
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7867 ATGGCGTCGAGAGTGGGTGCAAGCGCGCTGATCGGTTCCTGCTG 7818

408 ValIlePheSerPheCysSerLeuLeuLeuPheIleCysValGlyArgAs 424
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7817 CTGTGGCGCTCGCTGCTGCTTCTTACGCTCGCTGCTGCGGAT 7768

424 nValLeuThrLeuLeuPheIleAlaArgAlaPheIleSerGlyGlyP 441
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7767 GATTCGCCCTCGCTTCTCATCATGAGCTTCGCGCTGCTC...GCCACT 7721

441 heGlnAlaCysTyrValTyrThrProGluValTyrProThrAlaThrArg 457
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7720 GGGGTGCGCTTTATCGCTTTTACCGCGGAATTTACCCGACGCGCGG 7671
```

```
458 AlaLeuGlyLeuGlyThrCysSerGlyMetAlaArgValGlyAlaLeuI 474
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7670 GCGACCGCATGGCGCTCGCGGTCCATGGCAAGCTCGGGCGCTGCT 7621

474 eThrProPheIleAlaGlnValMetLeuGluSerSerValTyrLeuThrL 491
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7620 GCGCGCTCACCTGCTGTTATGCTATGCTGCGAGGTTTGGGTGCGGA 7571

491 euAlaValTyrSerGlyCysCysLeuLeuAlaAlaLeuAlaSerCysPhe 507
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7570 TTGGCATCTTTGCGCGGTACTGCTGCTGCGGCCATTCGCGCGACCATG 7521

508 LeuProIleGluThrLys 513
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7520 ATCAATGCGGAACACGG 7503

seq_name: gb_ba:AE007266

seq_documentation_block: 11856 bp DNA BCT 15-AUG-2001
LOCUS AE007266
DEFINITION Sinorhizobium meliloti plasmid pSymA section 72 of 121 of the
complete plasmid sequence.
ACCESSION AE007266 AE006469
VERSION AE007266.1 GI:14523915
KEYWORDS
SOURCE Sinorhizobium meliloti.
ORGANISM Sinorhizobium meliloti
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group:
Rhizobiaceae; Sinorhizobium.
REFERENCE 1 (bases 1 to 11856)
AUTHORS Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Abola,A.P.,
Barloy-Hubler,F., Bowser,L., Capela,D., Galibert,F., Gouzy,J.,
Gurjal,M., Hong,A., Huizar,L., Hyman,R.W., Kahn,D., Kahn,M.L.,
Kelman,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R.,
Wells,D.H., Yeh,K.-C., Davis,R.W., Federspiel,N.A. and Long,S.R.
Nucleotide sequence and predicted functions of the entire
Sinorhizobium meliloti pSymA megaplasmid
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9883-9888 (2001)
PUBMED 11481432
REFERENCE 2 (bases 1 to 11856)
AUTHORS Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Abola,A.P.,
Barloy-Hubler,F., Bowser,L., Capela,D., Galibert,F., Gouzy,J.,
Gurjal,M., Hong,A., Huizar,L., Hyman,R.W., Kahn,D., Kahn,M.L.,
Kelman,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R.,
Wells,D.H., Yeh,K.-C., Davis,R.W., Federspiel,N.A. and Long,S.R.
Direct Submission
JOURNAL Submitted (29-MAR-2001) Biological Sciences, Stanford University,
371 Serra Mall, Stanford, CA 94305, USA
FEATURES
source
location/Qualifiers
1..11856
/organism="Sinorhizobium meliloti"
/strain="1021"
/db_xref="taxon:382"
/plasmid="pSymA"
complement(116..1468)
/gene="Sma1447"
complement(116..1468)
/gene="Sma1447"
/notes="Glimmer prediction: global similarity to Benk, a
benzoate transport protein in P. putida; Pfam domain
homology to sugar and other transporters"
/codon_start=1
/transl_table=11
/product="putative transmembrane transport protein"
/protein_id="AAK63916"
/db_xref="GI:14523916"
/translation="MLIAAROLEKGRMTTITVDDALDRAGTGTYORRLAIFGLVNA
ADAMQVLAVGFTAASTAEGLTPQALQGTGLFFGLMFGAAGFGLADRIGRRLV
IATVACDAGLLSVFAQDTVLLLRFLTGAAGVGLPVDYAMAEFLFARNRRL
VMLEGFVAGTLLVLAANAASLAGADAWRYIFAVTAIPALIGVGLRFLVPESPLYL
LRIGKTSKAIIVDEILVYKMRGLAGASLPPPTASAGIFSDLRKSLMTLAIW
FLVSIYGVFTWMPRLAGEFGFVGLVVLVLAQIPGVLAAYGVKWKRRRT
LIGFCLLSALGCLLFVAAGTAMLIGVSLIMSALLGTWGLYAYTPELYTASRATG
```

```

MGAGAMARLGGLLAPSLMLVUQSGFLAIGFAGLLVAAVAFLIDAETRRVSLA
/
complement(1918. .3108)
/gene="Sma1450"
CDS
complement(1918. .3108)
/gene="Sma1450"
/function="Small molecule metabolism"
/note="glimmer prediction; global similarity to PhaA, a
thiolase from P. putida (GI:1857942); global similarity
to acetyl-CoA acetyltransferase from C. acetobutylicum and
A. eutrophus (GI:1174677); Pfam domain homology to
thiolase"
/codon_start=1
/transl_table=11
/product="Probable thiolase"
/protein_id="AAK65449.1"
/db_xref="GI:14523917"
/translation="MNKQDPVIVGQARTPLGSGFQGLKDLAADIAGAAIYDALKKA
GLAPDAVEMFGCVLTAGQGPAPQAALAGLPGVGATTVNMKCGSMKAAMLAH
DLIAKASASIVVAGMESMTNAPYLDDRAGYRIGHQKVLDMHFDLGLDAIDKGRH
MGSFADCAEQFTFRSAQDEYAIASLEKAKASADGSAEIEVLPSIASGKERTVN
LQEPQKARLDKIPLLKPAFRDGGITTAANASSISDGAALVLMRRSADKQIGPLA
VICGHATHADAPSLPTAPIGAIKALCRRIGWDIGEDLFEINEAFVVPMAAMRELG
LDAEVNHHGACALGHPIGASSARVIVTVNALRRRLRGLRGSIASVCIGGGEATAVV
FISG"
complement(3133. .3900)
/gene="Sma1452"
complement(3133. .3900)
/gene="Sma1452"
/function="Small molecule metabolism; Central intermediary
metabolism"
/note="glimmer prediction; global similarity to eukaryotic
3-hydroxyacyl-CoA type II (GI: 2569966); Pfam domain
similarity to short chain"
/codon_start=1
/transl_table=11
/product="Probable"
/protein_id="AAK65450.1"
/db_xref="GI:14523918"
/translation="MOLKSRFIVTGCASSGLGAAVTRMLAQRCATVGLGLDKPPAGEE
PAELGAARFRNADVTNEADATLAFARQEPGHVGHNCATGAPGKILGRSGPH
ALDSFARTVNLGTFNIRLAAEYMSQGEDADGGERVIVNTASIAAFDGGIGQAA
YAAKGGVAALTLPAARELARFGIRVVTAPGIDFPMWAGMPQVDQDALAASVPFP
RLGRAEYALVKHICENTMLNGEIVRLDGLARMAP"
3995. .5065
/gene="Sma1454"
3995. .5065
/gene="Sma1454"
/function="Miscellaneous; Not classified regulator"
/note="glimmer prediction; global similarity to OruR from
P. aeruginosa (GI: 1657970), which regulates ornithine
utilization; pfam domain similarity to N-terminal
helix-turn-helix motif"
/codon_start=1
/transl_table=11
/product="Putative transcription factor"
/protein_id="AAK65451.1"
/db_xref="GI:14523919"
/translation="MFNEKGPShwerrmIAPCFMDTLECLRHRGVDAQPLLAQAGLP
SIVTGPVSANQCAFHVAQAQMDDEFEGGASPMRAGSFALLCHALTSLATLEHALR
LRLRVLVENPHGELVWENLAQIVLKHQDQATRSFAYRTFWILHGVNCHLIGRR
LPIRWDFRCSAPAGTLYVLFPGVPDQRRTRIVDAEYDKLPPIDREALKHPL
RHAPANILVRYRHADGLSAAIQRLHATAPSAWPGFEAIAARMRIAPATLRRRLRQEG
QTVRSIKEDLRALAEALADGRTNVAQLAVELGFSFSAFHRAFRKWTGSKSPAQFRR
NATQTFDAESSSPKRKTQPOES"
complement(5047. .5565)
/gene="Sma1455"
complement(5047. .5565)
/gene="Sma1455"
/function="Miscellaneous; Unknown"
/note="glimmer prediction"
/codon_start=1
/transl_table=11

```

```

/product="Hypothetical protein"
/protein_id="AAK65452.1"
/db_xref="GI:14523920"
/translation="MKRTGTCYRAHDPRWAFKPSGDGAAIKGARFNPKGVKTYLAL
SMTAVKANQGFARHIDPCVLCSDIDCADIADLTAEGRAEHGVSLFEACMSWAGA
FAEGRRPASWAIYNRLHPRGTAGILVPSFAPGTAGDRNLVLTWGPDLPHRLDYDP
SGRLPKDOLSWG"
complement(5562. .5951)
/gene="Sma1456"
complement(5562. .5951)
/gene="Sma1456"
/function="Miscellaneous; Unknown"
/note="glimmer prediction"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK65453.1"
/db_xref="GI:14523921"
/translation="MAOAKIQSDGGLPILSYMDKGGKIAYOVAVGFQMSKTLQAE
AGLARETLVPERSRGTONTLRMLEIISRTWDAGKEQAMWYRAQPLPAFGGR
TABALVKDKAAAVRDYLDHMLGGFA"
complement(6059. .7063)
/gene="Sma1459"
complement(6059. .7063)
/gene="Sma1459"
/function="Small molecule metabolism"
/note="glimmer prediction; global similarity to
malate dehydrogenase from M. fervidus (GI: 80950)"
/codon_start=1
/transl_table=11
/product="Putative"
/protein_id="AAK65454.1"
/db_xref="GI:14523922"
/translation="MTPVETSMEEIFAVCLEALTLHCAAAETARAVAGAITRAEADCN
RVCGLYLPFCRHILAGIKVDEAVPEVTGVTVDARSQFAHPAIAAGTPTALIDL
AKOAGLAAMAVRNSYNCLALGHHVRPLADAGLIGCVSNAPASVAPPGATRALFGTNP
LAFVPSKEGAPTIVVDQMSAVTKTEMLIRDRDGEAIPIGWADGNGQTITDARTGL
EGSLPAGGRKGANIALLEVIAAALTGSLASTEASAFNEEGGPHVGGFLAIDPD
HFAAGHFSEAMDNLVASHDAAGVRLPGHFGKQPCVCDADLKKKAVILSKSKNRQKHG"
complement(7060. .8184)
/gene="Sma1461"
complement(7060. .8184)
/gene="Sma1461"
/EC_number="5.5.1.1"
/function="Small molecule metabolism; Central intermediary
metabolism; amino acids"
/note="glimmer prediction; global similarity to muconate
cycloisomerase in P. putida (GI: 77758); Pfam domain
similarity to mandelate racemase / muconate lactonizing
enzyme family"
/codon_start=1
/transl_table=11
/product="Putative muconate cycloisomerase"
/protein_id="AAK65455.1"
/db_xref="GI:14523923"
/translation="MKRISNVRRLVPLKQPHWSYGRISFAVNLIEADGTV
GIGETVAPDQTGTAAIILYKALHVLGSHSPHVAPLIARFHEYLGHGANIMRAAQ
IFSGIDAMMDLQGLAGLVHQLGAGMKAVGYFFLQGETAEELARDAAVGHAQG
EFTVYLKVGEGKLDLEITAAVRGEIDARLDDANEGSVHDAINMKRLEKYEIEF
IEQPTVSWSPANAHVREKVGIPVADQAAFTLVDDVEICRQRAADMICIGPREIGGI
QPMKAAVAEAGLKICIHSSFTTGITTCAEHHIGIAIPNLDGNGIIMQLVQEDIV
SSDPLTPKNWLDQAFKPGGLGFPQLAEDLVAGEGGRYAASR"
complement(8311. .9201)
/gene="Sma1462"
complement(8311. .9201)
/gene="Sma1462"
/function="Cell processes; Transport of small molecules"
/note="glimmer prediction; global similarity to
choline-binding protein precursor from B. subtilis
(GI:7404430) and several putative ABC transporter
proteins; adjacent ORFs encode other putative members of
the choline transport system"

```

alignment_scores:
 Quality: 540.50 Length: 458
 Ratio: 1.917 Gaps: 9
 Percent Similarity: 61.572 Percent Identity: 30.131

alignment_block:
 US-09-911-667A-2 x AE007266/rev ..

Align seg 1/1 to reverse of: AE007266 from: 1 to: 11856

```

67 ThrPheMetValGluAspAlaValGluAlaIleGlyPheGlyLysPheG1 83
1423 ACAATACAGTAGACATCGCTCGATCGCGCGCGGACACTTACCA 1374
83 nTrpLysLeuSerValLeuThrGlyLeuAlaTrpMetAlaAspAlaMetG 100
1373 GCGCAGACTGATGGCATATTCGGTCTGGTGGCTGCCGATGCCATGC 1324
100 LuMetMetIleLeuSerIleLeuAlaProGlnLeuHisCysGluTrpArg 116
1323 AGGTGCTGGCGTCTCGCTTCACCGCGCATCGATCGCGCGCACTTTCGCG 1274
117 LeuProSerTrpGlnValAlaLeuLeuThrSerValValPheValGlyMe 133
1273 CTGACAGTTCGCGCAGCACTTCAGACGGCGCATCTTCTTCGGAAT 1224
133 tMetSerSerThrLeuTrpGlyAsnIleSerAspGlnTrpGlyArgL 150
1223 GCTGTTCCGCGCGCGGTTTCGCGAGGCTCGCGACCGGATCGCGCGCC 1174
150 yThrGlyLeuLysIleSerValLeuTrpThrLeuTrpGlyIleLeu 166
1173 GCGCGCTTCGATCGCCACCGTCTCGCTCGGAGCGCTCTCGGTCTGCTG 1124
167 SerAlaPheAlaProValTrpSerTrpIleLeuValLeuArgGlyLeuVa 183
1123 TCGGTCTTCGCGCAGCACTTACCGTTTTCGCTTCGCTCGGTTCTCAC 1074
183 lGlyPheGlyIleGlyGly...ValProGlnSerValThrLeuTrpAlaG 199
1073 GCGCGCAGCTGTCGCGCGGACACTCGCGTCACTATCGGATGATCGCGG 1024
199 LuPheLeuProMetLysAlaArgAlaLysCysIleLeuLeuIleGluVal 215
1023 AGTTCTCGCGCGCGGAAACCGCGCGCTGGTGTGTCATGCGGAAGGC 974
216 PheTrpAlaIleGlyThrValPheGluValValLeuAlaValPheValMe 232
973 TTTGGGCTGTCGCGACGCTCAFC...GTGGGCTCGCGCGCTGGGCTGC 927
232 tProSerLeuGly.....TrpArgTrpLeuLeuIleLeuSerA 245
926 GAGCCTCGCGCGGCTCGCGATGCTCGCGGATACATCTTCGAGTACGCG 877
245 lValProLeuLeuLeuPheAlaValLeuCysPheTrpLeuProGluSer 261
876 CAATCCGCGCAGTGTATGCGCGGCTCGGTTCCTGCGCGGATCG 827
262 AlaArgTrpAspValLeuSerGlyAsnGlnGluLysAlaIleAlaThrLe 278
826 CCGCTTTATCTGCTCGCGCTGGGAAAGACGTCGAGCGGAGGCCATCGT 777
278 uLysArgIleAlaThrGlnAsnGlyAlaProMetProLeuGlyLysLeu 295
776 CGATGAGATCTCTGCTGTCACAGGC...AAATGCGACTTGGCGCGCGG 730
295 lLeuSerArgGlnGluAspArgGlyLysMetArgAspLeuPheTrpPro 311
729 CATCGCTGGTTCGCGCGCGCGCGCGGATCGCGGTATTTTTCGCGG 680
312 HisPheArgTrpThrThrLeuLeuLeuTrpPheIleTrpPheSerAsnAl 328

```

seq_name: gb_ba:AE001804

seq_documentation_block:

LOCUS AE001804 17641 bp DNA BCT 02-JUN-1999

DEFINITION Thermotoga maritima section 116 of 136 of the complete genome.

ACCESSION AE001804 AE000512

VERSION AE001804.1 GI:4982160

KEYWORDS Thermotoga maritima.

SOURCE Thermotoga maritima

ORGANISM Thermotoga maritima

REFERENCE 1 (bases 1 to 17641)

AUTHORS Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J.,

Hart, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A.,

McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M.,

Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A.,

Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D.,

Eisen, J.A., Fraser, C.M. et al.

Evidence for lateral gene transfer between Archaea and bacteria

from genome sequence of Thermotoga maritima

Nature 399 (6734), 323-329 (1999)

JOURNAL 92987316

MEDLINE 92987316

REFERENCE 2 (bases 1 to 17641)

AUTHORS

Nelson, K.E., Clayton, R.A., Gill, S.R., Winn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A., McDonald, L., Uitterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M., Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A., Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D., White, O., Salzberg, S.L., Smith, H.O., Venter, J.C. and Fraser, C.M.
Direct Submission
Submitted (01-JUN-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

TITLE
JOURNAL

FEATURES
source

Location/Qualifiers
1. .17641
/organism="Thermotoga maritima"
/db_xref="taxon:2336"
146. .580
/gene="TM1590"
146. .580
/gene="TM1590"
/note="similar to GB:AE000657 percent identity: 74.38;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="translational initiation factor IF-3"
/protein_id="AAD36657.1"
/db_xref="GI:4982161"
/translation="MTGVPTKALEAREKGLDLVLVAPNPPVAKIMDYGKYKYQ
LTKQKRNKKPVQMKQKFRUKIDEDHQTQVKHHLRFEEDGHKVRVVMFVIGREMM
FAEKGKELLERVTKDTELTATVESPMEGRDMMVVLKPKNS"
595. .792
/gene="TM1591"
595. .792
/gene="TM1591"
/note="similar to SP:P55874 PID:1770008 GB:AL009126
percent identity: 65.62; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="ribosomal protein L35"
/protein_id="AAD36658.1"
/db_xref="GI:4982162"
/translation="MPKVTNRSAAKRFRTKNGKIMRNHAYRSHKTGKKRRNALRAL
RKDDVVSSADKNVRLRLGKK"
806. .1162
/gene="TM1592"
806. .1162
/gene="TM1592"
/note="similar to SP:P55873 PID:1770009 GB:AL009126
percent identity: 69.57; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="ribosomal protein L20"
/protein_id="AAD36659.1"
/db_xref="GI:4982163"
/translation="MRVKRAVHAKKKKKYKKAAGYRGALSHRYKLAKOMYVKKWY
SYVGRKKRDMRKLMTIRITNIAARNEGLKYSLEIHLGLAGVSNKKMLSELAVNDP
EAFKEYVIAKEALAS"
1173. .1916
/gene="TM1593"
1173. .1916
/gene="TM1593"
/note="similar to GB:Pyro.h percent identity: 72.27;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAD36660.1"
/db_xref="GI:4982164"
/translation="MEKTKKIAYMSGKGVGKTTAVNLAVALAAEGYQVGLDLDIH
GNVQRMGLVSLPPSEGEKIVPAKYGSLKVSFLAMIQEGAPVWIRGLKHAIEQL
TRVWEGDLYLIDLPDGTGDEALSTFQIKPDATIVSTPQKVGDDVRRRAINFVK
RLSKGILGLVNNMSYLVCPNCGEKIVVFGKTEKLAEEFGIPLIARIPMDPEVSL
DEGRPAVYKRGTVIEEEKVIEKVLSL"
complement(1882. .5502)

gene

146. .580

CDS

146. .580

gene

595. .792

CDS

595. .792

gene

806. .1162

CDS

806. .1162

gene

1173. .1916

CDS

1173. .1916

CDS

/gene="TM1594"
complement(1882. .5502)
/gene="TM1594"
/note="similar to GB:AE000657 percent identity: 60.77;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein, GGDEF domain"
/protein_id="AAD36661.1"
/db_xref="GI:4982165"
/translation="MMKVRKFKFLRETYMGDEFLEDDGEDQVLKIVRAPVDRSFFFN
EYAKLRNLNPNVLLPEKLIKIDSGKFLFPYVYHNLALENISDQVAKQLIKLFLFLS
KVGVTIPALGMDLILVNDGVFLIPALISNIPDVGCVFSPERSRATTEVCRFLKI
HGLEPHLETRESFDSREIRIPYIHRKEELIKBDIEIAKPFPIILITGEORVGTK
IASVLVDGREQGYMVHQITSLDLRWITDASDELDTLLKDCGKKVILIDDLFGS
DLISFLFEESGLTITKIIVLTSTKAFQFHHKVRFLSPFTVEETRIFLERAIGKISE
DOVKLIHSLKGLPGYVVELLKFNFKNLNRENIVGIFKPLLELDSPEIRLSVLGOK
FTGTELVLEKITGDKYHDTLMSAYDSGVITTEEGYRFALREFMYYNKLSNKKK
NTHKELFENLPUDLIAHLSLEDPRLKFTFYVLRVYRKHFWDEKTRSLI EYLRLEEE
FEKPYVSTESLKMKLI FRIDPLGVKEKFNHFRFRKTLRNPVDESILQKDSLSYDIY
NIULSRPHIRAGKAKPOELLNTIKRELKKNFSTRERLYLKAHLIDYDLSSFDREA
LKEMNIATSEGFLDIQVMYRGALGVLSRTRAMSNYYFHHLSLSKIDPSLSDVES
NLTSLLYEGKITFLVQLERLRKQARLFEDMPILSYTYFLEGLYTHKKRDFOKAEV
FRTLELEKHGIERALRGLVNNYLFSGDVESAKSLLEKDEFEFDFRGNFLRLVYL
AKDSSLIKWKELETPOKFFNEEYAVVTERLAKLDPEGFVEFLELELRNENSS
NLTALVYTFYKFYSALGNPKAKRYLRRAIFVYNLIGLREVSVPKVEPTEKTOIEEK
KAPYLLGLFIEKEKESOMMEFASARLSEVIPYEVFSIRIERTTKVMEYSTPI
TPMEKDFLEISPFRTVMGLSKMDHMIVGVETNLECKDEKTAWEIVETLEOPGNILT
TILRERLYRDRSMKDLTGTVLHSMYFMERLEEAYKSSRYKSPLSIIMCDADDFKIN
DOFGHVAGDKTLGWLGRKMKSVLRKSDLVGRYGGEEFI IALPGTSLEEAIVAEKLK
AVMEDPENTYHITLSFGVABYKNGEDPFETIKRADEALYLAKILGKNSVVVTEKLSRR
SS"
complement(5450. .6451)
/gene="TM1595"
complement(5450. .6451)
/gene="TM1595"
/note="similar to GB:AE000657 percent identity: 58.22;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAD36662.1"
/db_xref="GI:4982166"
/translation="MDEIVKVLQSHQDRILVVGHHMPDGCSSVLSLTGLKGLKREV
KAAVDYKIPVFEKFPYIDRIEENPNDFPELLVVVDASSPDRIGKFDOLLDKPVSVI
DHRTNTNFGNMWVDPSPAATQMIPIRINKALGVEDSNLATNLNYLGIDTGYGFRH
SNADRVFEDAYKLKMGADAHFVAKELLENKREFOFKLFAEVLRLQLLENGKIAYS
YTDYTLRHNCIDEDSAGFVGLSRIGVEAVLMEFPFGKIHVSMRSKDMFNVE
VAFELGGGHPRAAGVTFEGKKIEEVIIPRVINHLKFKREGVESESEKIPEGDVLGG"
complement(6452. .7249)
/gene="TM1596"
complement(6452. .7249)
/gene="TM1596"
/note="similar to SP:P46354 PID:944944 PID:1303996
GB:AL009126 percent identity: 68.54; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="purine nucleoside phosphorylase"
/protein_id="AAD36663.1"
/db_xref="GI:4982167"
/translation="MMKKIEARTFISERTNLSPDILLILGSGFGPPIEKVEDPVIID
YKDIPIHPQPTVEHSGKLVFGRIISDKPVMIMAGRPHFLYEGHDPAIVAFVPIYLKAYG
VKGVVPTNAAGAINPEFKGEIILVRDIINFMFRNPLRGPNDEKIFRPFDDMSVDDP
EWARKQERLSLKEGYVIGLVGPSYETPAETIRVFEKLGADLVGMSTVPIAAKHGCL
KVVFSCVTNNAAGIITHGRLSHEEVVTRTKMAQCKIEKALTAVEVE"
7299. .8363
/gene="TM1597"
7299. .8363
/gene="TM1597"
/note="similar to percent identity: 0.00; identified by
sequence similarity; putative"
/codon_start=1

gene

complement(5450. .6451)

CDS

complement(5450. .6451)

gene

complement(6452. .7249)

CDS

complement(6452. .7249)

gene

complement(6452. .7249)

CDS

complement(6452. .7249)

Tue Mar 12 15:54:23 2002

us-09-911-667a-2.rge

```
/transl_table=11
/product="hypothetical protein"
/protein_id="AAD36664.1"
/db_xref="GI:4982168"
/transl_table="MYPRLIILNKEIEENARKVEMASRRGIIVGVTKVTLGDPDRF
AETLRKAGICIGESIRIRNLRMKKAGIEGPFMLRLPMSEIIEVDKHFDFIMVSDP
DVAKKVDLSREMRKRVKLIIMIDVGLREGVWFKEAVEIAOGRGANIVGIGTINFGC
YGGIIPREKFEILLDIKELKXNHSFNIEVSGGCTPALALENGEIPGGINOLRIG
EATVLRGDTNNRVIDWLSONTFLIEAEVIEVKEKPSVPLGRGLDVFGRKVDVDRG
IRKRAICAGEQIDISRGILPDKGVLEVLIHSSDHLVLDVDFGDKVKVGDVFRFRMTY
SCLLKAMTSPFVEKVEPSI"
8396..8965
/gene="TM1598"
CDS
8396..8965
/gene="TM1598"
/notes="similar to SP:P34086 PID:1045628 PID:987648
GB:000096 PID:1788926 percent identity: 61.67; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="RNA polymerase sigma-E factor"
/protein_id="AAD36665.1"
/db_xref="GI:4982169"
/transl_table="MNEKKLVEALKKEDWAYEVLYREFAPKIGSIKSLNTUDDVDD
VVDVFLRIKFKIGKFRGDSKSLTWIYRIAVNVCKDYLOKTKRNEFLDQFQEKNEE
EIIYIQEAKENLVDEVLRSITAEVORALEKLSPEDRLLIKLRDIDGLSYEEIASVLO
KPVGTVKRSLHYARKRLGNLLKEGTVER"
8946..9308
/gene="TM1599"
CDS
8946..9308
/gene="TM1599"
/notes="similar to percent identity: 0.00; identified by
sequence similarity; putative"
/codon_start=1
```

```
alignment_scores:
  Quality: 527.50      Length: 523
  Ratio: 1.758        Gaps: 15
  Percent Similarity: 57.361 Percent Identity: 29.063

alignment_block:
US-09-911-667A-2 x AE001804 ..
Align seg 1/1 to: AE001804 from: 1 to: 17641
13 ValVallysPheArgThrGlyGluSerAlaArgSerGluAsp...As 28
12113 GTCTTGAATTTGAACTTGATGGATGCGCAAGACACAACTCTCTT 12162
28 pThrAlaSerGlyGluHisGluValGlnIleGluGlyValHisValGlyL 45
12163 AACTCTCTCTGA.....GGAGTTTCAT.....C 12185
45 euGluAlaValGluLeuAspGlyAlaAlaValProLysGluPheAla 61
12186 TTCACACATAGAGCCCGGATGAACAACATGGAAAGAAATCATGAGA 12235
62 AsnProThrAspThrPheMetValGluAspAla..... 73
12236 GAGCTGAAGAAAAAGGTTTCTCATAGAGGAGGATAAAAGATGAGAA 12285
74 .....ValGluAlaIleGlyPheGlyLysPheGlnTrpLysLeuS 87
12286 AGACGAAATAGTGTAAATAACGTGGACAGAAAACTCAGAGAAGGTTT 12335
87 erValLeuThrGlyLeuAlaTrpMetAlaAspAlaMetGluMetMetile 103
12336 TCATCTTCATCTTATCGCTGGATGTTTCGATGCTGCGGTTTATGCTT 12385
104 LeuSerIleLeuAlaProGlnLeuHisCysGluTrpArgLeuProSerTr 120
12386 CTTCTTTTGTCTCTCTACGTGATATAAAGAAATGGAATCTCACTCTAC 12435
```

```
120 pGlnValAlaLeuLeuThrSerValValPheValGlyMetMetSerS 137
12436 TCAGGGAGCTACCATAGCAAGTGTACCTTCCTTGAATGCTTTTGGAG 12485
137 erThrLeuTrpGlyAsnIleSerAspGlnTrpGlyArgLysThrGlyLeu 153
12486 CGCTTCTCGGGGTTCTGCTGACCTTCTTGAAGGAAAGTTTCCAAT 12535
154 LysIleSerValLeuTrpThrLeuTyrTyrGlyIleLeuSerAlaPheAl 170
12536 CTGTTCTTTTCATCATCAAGCATCACCTTCTCTGTCGGCTTTTC 12585
170 aproValTyrSerTrpIleLeuValLeuArgGlyLeuValGlyPheGlyI 187
12586 CTCATCGTTCGAAACTCTTTAGTCTCTGAGAGGCTTTCGGGTTTGGGT 12635
187 leGlyGlyValProGlnSerValThrLeuTyr...AlaGluPheLeuPro 202
12636 ATGAGGCTTCATGCTTCGTTCAACGGGTATCTTGAGAAATTCACGAGT 12685
203 MetLysAlaArgAlaLysCysIleLeuLeuLeuGluValPheTrpAlaI 219
12686 ATCAGACTCAGGGGAAGGTATCTTGTGCTTCTGAGTCTGAGCTGGCGGT 12735
219 eGlyThrValPheGluValValLeuAlaValPheValMetProSerLeu 236
12736 TGGGAGTATAGTATAGGACTCTTCGCAGTGAATGCTTCTCCGAAC... 12781
236 lyTrpArgTrpLeuLeuLeuSerAlaValProLeuLeuLeuPheAla 252
12782 ..TGCAGATGGTCTCTTGGATCTTCTCCATCGGG...TACCTTTTGT 12826
253 ValLeuCysPheTrpLeuProGluSerAlaArgTyrAspValLeuSerG 269
12827 CCGGTGTTCTTAGGATGCCGAAACCCAAAGTACGCGTTTTCGAGGG 12876
269 yAsnGlnGluLysAlaIleAlaThrLeu...LysArgIleAlaThrGluA 285
12877 TGGAAAAAGAGCGCTGGAGAGTCTCTGGGTAAAGAGTGGAGAGAGAG 12926
285 snGlyAlaPro.....MetProLeuGlyLysLeuIleLeuSer 297
12927 TGGAGCTTCCAAAAAAGAGAAAGTTCCCATCTCTGCTCTTTG..... 12970
298 ArgGlnGluAspArgGlyLysMetArgAspLeuPheThrProHisPheAr 314
12971 .....AAGAGGGAACATTTGAAAGAC..... 12991
314 gTrpThrThrLeuLeuTrpPheIleTrpPheSerAsnAlaPheSerT 331
12992 .....ACAGTTGTGATCTGGATAGCGTGGTTCGTTGTGAGCTTCGTCT 13034
331 yTrpGlyLeuValLeuLeuThrThrGluLeuPheGlnAlaGlyAspVal 347
13035 ATTACGCTCTGTTCACCTGGCTCCAGGATCTCTCTCTCG..... 13075
348 CysGlyIleSerSerArgLysLysAlaValGluAlaLysCysSerLeuAl 364
13076 CTGGGTGTGAGTGTGTCAAAATCTTC..... 13102
364 acysGluTyrLeuSerGluGluAspTyrMetAspLeuLeuTrpThr... 380
13103 .....TGGTTCACAT.....TGGTTCACAT 13112
381 .....LeuSerGluPheProGlyValLeuValThrLeuTrpIle 393
13113 TCATCATGATGGTGGCAGCTTCCCGGGTACCTTTCGGCAGCCTACTTC 13162
394 IleAspArgLeuGlyArgLysLysThrMetAlaLeuCysPheValIlePh 410
13163 ATAGAAAAATGGGGAAGAAAGGCTTCTTGTGTTTACTTTCATAGGAAC 13212
410 eSerPheCysSerLeuLeuLeuPheIleCysValGlyArgAsnValLeu 427
```



```
/gene="ZK637.4"
/notes="predicted using Genefinder"
/codon_start=1
/protein_id="CAA77451.1"
/db_xref="GI:3881666"
/db_xref="SWISS-PROT:P30637"
/translation="MKSNPXYFLMNDVQSKYSPKYPVNSLKERILHFLDYIAPL
KLYLLSYPPDCLWNRKLRKASGVQVTPSPSEPHIDRLIHISOKQPSF"
join(11124..11557,11606..11864,12073..12282,12411..12536)
/gene="ZK637.5"
join(11124..11557,11606..11864,12073..12282,12411..12536)
/gene="ZK637.5"
/notes="predicted using Genefinder"
Ar5A homologue
cDNA EST yk62f10.5 comes from this gene
cDNA EST yk62f10.3 comes from this gene
cDNA EST yk140f3.3 comes from this gene
cDNA EST yk176c11.3 comes from this gene
cDNA EST yk140f3.5 comes from this gene
cDNA EST yk176c11.5 comes from this gene
cDNA EST yk221f9.3 comes from this gene
cDNA EST yk221f9.5 comes from this gene
cDNA EST yk267c10.5 comes from this gene
cDNA EST yk507e12.3 comes from this gene
cDNA EST yk191h10.5 comes from this gene
cDNA EST yk507e12.5 comes from this gene"
/codon_start=1
/protein_id="CAA77452.1"
/db_xref="GI:3881667"
/db_xref="SWISS-PROT:P30632"
/translation="MSDQLEASIKNILEQTLKWIIVGGKGVGKTKTCSLAAQLSK
VREVLITSDPAHNTDAFQSQRTPTTLVEGKFLFAMEIDSNPNGEVEMGNIEE
MLQNAQEGSGFSGKDFLOSGFAGLPGIDEAMSGEMIKLIDSLDFVVFDTA
PTGHTLRLLQFPTLLEKFTKILSLQGMFGPMNFGMCGGSMNEMIEKMTITL
ESVKMNAQKPDNCTTFCVCIAEFLSYETELRLOELSKQGITDHNIIIVNOLLFPD
TDANGTVSKRCKASQALQSKYLTIDILEYDFHVVKLPILLEAEVGRGPAILQFSERM
VDPEANKN"
complement(join(16630..16827,17583..17732,17779..17883,
17962..18372,18848..19006,19051..19653,19694..19826,
19876..19994,20104..20160))
/gene="ZK637.7b"
complement(join(16630..16827,17583..17732,17779..17883,
17962..18372,18848..19006,19051..19647,19694..19826,
19876..19994,20104..20160))
/gene="ZK637.7a"
complement(join(16630..16827,17583..17732,17779..17883,
17962..18372,18848..19006,19051..19647,19694..19826,
19876..19994,20104..20160))
/gene="ZK637.7a"
/notes="predicted using Genefinder"
cDNA EST yk34b2.3 comes from this gene
cDNA EST yk24e1.3 comes from this gene
cDNA EST yk12h1.3 comes from this gene
cDNA EST yk24e1.5 comes from this gene
cDNA EST yk12h1.5 comes from this gene
cDNA EST yk34b2.5 comes from this gene
cDNA EST yk353h6.3 comes from this gene
cDNA EST yk353h6.5 comes from this gene
cDNA EST yk270e5.3 comes from this gene
cDNA EST yk270e5.5 comes from this gene
cDNA EST yk224d7.3 comes from this gene
cDNA EST yk224d7.5 comes from this gene
cDNA EST yk212f5.3 comes from this gene
cDNA EST yk212f5.5 comes from this gene
cDNA EST yk220e8.3 comes from this gene
cDNA EST yk220e8.5 comes from this gene
cDNA EST yk207e10.5 comes from this gene
cDNA EST yk296g5.5 comes from this gene
cDNA EST yk350b7.5 comes from this gene
cDNA EST yk430g11.5 comes from this gene
cDNA EST yk561d3.3 comes from this gene"
/codon_start=1
```

```
/protein_id="CAA77454.2"
/db_xref="GI:14530698"
/db_xref="SWISS-PROT:P30630"
/translation="MSSAVRSPRKKASDTSDDRTSSPYSLKRETSKVPSPRYNEELY
LSPRSIRKGTGSPKSPAKRLNGRSDPSVNSLTLMLAKAALDYESSSCALEY
LPEKERRPRRALALSPAPSDNLLAKOLEMHOHLVAGLDLDLNDPANMTNEAVE
HRDTQFFNNFSTDOQRSAMMKOFKTYTKNOTSDVSTFMRANTKLYNLLRYKKARQW
VACEFYSAIDQIFKEENEFAITRESFNLKNWNI.TRIEWSIRKLLGKPRRCSKY
FTEERMYLEERMKIRSVVEGSLNDPSLDLKLPAKLPMPVGVGNRVFARIRNPYD
GIYSGIIDAIVPKGRIIFDKPDIPTPLVSDTEILDGKLDLLSIAFYTEQANSKLPS
GVPPFAAVARDSSPHLVRDVLVSRKIERSGGFLMGPNDRNLNGKNAEMVGNFPLKFL
```

```
alignment_scores:
  Quality: 511.50      Length: 458
  Ratio: 2.483        Gaps: 13
  Percent Similarity: 44.978  Percent Identity: 31.659

alignment_block:
US-09-911-667A-2 x CEZK637/rev ..

Align seg 1/1 to reverse of: CEZK637 from: 1 to: 40699

42 HisValGly.....LeuGluAlaValGluLeuAspGlyAl 54
|||||.....:|||||
1367 CATGTCGGAGATGATTTTCCAGTTCGATACCAACCTCGATGATCGTAC 1318
|||||.....:|||||

54 alaValProLysGluPheAlaAsnProThrAsp.....66
:|||||.....:|||||
1317 T.....GAACCTGGAGAACCTACAGATCAGAGATCACCAGATA 1280

67 .....ThrPheMetValGluAspAlaValGluAlaIleGlyPheGly 80
|||||.....:|||||
1279 GTGAAAAGACGTTTACAGTTGATGAAGCTGTTGAAGCATTTGGATTTGA 1230
|||||.....:|||||

81 LysPheGlnTrpLysLeuSerValLeuThrGlyLeuAlaTrp.....94
:|||||.....:|||||
1229 ACATTTACGCTCAAACTGCTCAATTTTGACAGGAATGGCATGGCTGATTT 1180

94 .....94

1179 CGAGGAGGATTTTCAAAAAAGAGTGTTCAATCCAAATGAGTTTATATATAC 1130

94 .....94

1129 AATATATTTTCACAATTTGAATTTGAAATTTGGAATTTTACCAAAAT 1080

94 .....94

1079 TAGATTGAAACTTTTGAAATAAAAAAACATTTTAGAGCAGATTTCTG 1030

94 .....94

1029 TTGCATGACTGTAATGAAACATTTCAAAATTTATTTTAAAAATATCAAA 980

94 .....94

979 AAAATAGTCAGGAAAGACCAAGAAATTTTAAACACCTCAGAAAAAATAA 930

94 .....94

929 GAATCTTCAATATGATTTTAAAAAAATTTTATTAACACACCTTTCCGCA 880

95 .....Met..AlaAspAlaMetGluMetIleLeuSerIleLeuAl 108
|||||.....:|||||
879 GAATTTTCAGATGCCCGCCCAATGGAATCATGTTTATTTATCATTTGATTC 830
|||||.....:|||||

108 aProGlnLeuHisCysGluTrpArgLeuProSerTrpGln.Val.....122
:|||||.....:|||||
829 TCCAGCTCTGGCATGTGAATGGGAATATATCGTCAGTGCACAGGTAAGTT 780

123 .....AlaLeuLeuThrSerVa 128
|||||.....:|||||

779 TCAATCAAGCAAAAAAATTTGTTGAGGCTCTTGTGACGACGCTG 730
```

```

128 lValPheValGlyMetMetSerSerThrLeuTrpGlyAsnIleSerA 145
||||| ||||||| ||||||| ||||||| ||||||| |||||||
729 CGTGTTCAGTGGATGATGTTCTCAAGTACATTTTGGGAAAAATATGTCG 680
145 spGlnTyrGlyArgLys..... 150
||||| |||||||
679 ATCGATTGGTCGACGAAAGCTAGTTGAAATATATAAATGATATAAATGA 630
151 .....ThrGlyLeuLysIleSerValLeuTrpThrLeuTy 162
||||| ||||||| ||||||| ||||||| ||||||| |||||||
629 AATCTCTAAAAATTTTCAGGACATAACATTTTCAACACACTAGTTCTGTGA 580
162 rTyr.GlyIleLeuSerAlaPheAlaProValTyrSerTrpIleLeuVal 178
||||| ||||||| ||||||| ||||||| ||||||| |||||||
579 TCATGGGTGTCATTTCTCGAATGCACCTCATTTTACGTTCTCTATTT 530
179 LeuArgGlyLeuValGlyPheGlyIleGlyValProGln..... 192
||||| ||||||| ||||||| ||||||| ||||||| |||||||
529 TTCCGTGGACTTACTGGATTTGGTATCGAGGTGTACCCAGTCGTTAGT 480
193 .....S 193
479 TTTTAAATCAGAAATATCCAAATCCCTATTTTAAAAAATTCATCTTTC 430
193 er.ValThrLeuTyrAlaGluPheLeuProMetLysAlaArgAlaLysCy 209
||||| ||||||| ||||||| ||||||| ||||||| |||||||
429 AGTGTCACCTTATACGTCGAGTTCTTCCAACTGCTCAACGTGCCAAATG 380
209 sIleLeuLeuIleGluValPheTrpAlaIleGlyThrValPheGluValV 226
||||| ||||||| ||||||| ||||||| ||||||| |||||||
379 TGTAGTTTTTGTAGAGATCAATTTTGGGCTATCGGAGCAGTTTTCGAAGCTC 330
226 aIleuAlaValPheValMetProSerLeuGlyTrpArgTrpLeuLeuIle 242
||||| ||||||| ||||||| ||||||| ||||||| |||||||
329 TTCCTGCTTATTTGTGATGGAATCATTTGGTTGGAGACCTTATATGTTTC 280
243 LeuSerAlaValProLeuLeuLeuLeuAlaVal.LeuCysPhe..... 256
||||| ||||||| ||||||| ||||||| ||||||| |||||||
279 CTTTCTTCTACACTCTCTGGAATTTTTCAGTGGCTTCTTTTGTAAAGTTT 230
256 ..... 256
229 TATCAAAAAATTCAGCTTAGAATTCAAAAATGCTGAATTCAAAGCTAG 180
257 .....TrpLeuPr 259
179 TATGCAAGTATAGTATACAAAGTAAACTTCAAAAATGTTTTCAGTGGCTCC 130
259 oGluSerAlaArgTyrAspValLeuSerGlyAsnGlnGluLysAlaIleA 276
||||| ||||||| ||||||| ||||||| ||||||| |||||||
129 AGAATCTGCTCGTTTGGACATGCTTCTGGACATCCCGAACAGCTCTAG 80
276 lathrLeuLysArgIleAlaThrGluAsnGlyAlaProMetProLeuGly 292
||||| ||||||| ||||||| ||||||| ||||||| |||||||
79 AAACACTTCAAGCAGCAGCAGGAGAGTGAATAGAGTACAGTTGCCAAGTGA 30
293 LysLeuIleIleSer 297
||||| ||||||| ||||||| ||||||| ||||||| |||||||
29 AGATTAGTGTCTTCA 15
seq_name: gb_pat:AX127153

```

```

seq_documentation_block:
LOCUS AX127153 309400 bp DNA PAT 11-MAY-2001
DEFINITION Sequence 7069 from Patent EP1108790.
ACCESSION AX127153 AX114121
VERSION AX127153.1 GI:14041141
KEYWORDS
SOURCE Corynebacterium glutamicum.
ORGANISM Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriaceae; Corynebacteriaceae;

```

```

Corynebacterium
1 (bases 1 to 309400)
Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K.,
Yokoi,H., Tateishi,N., Senoh,A., Ikeda,M. and Ozaki,A.
TITLE Novel polynucleotides
JOURNAL Patent: EP 1108790-A 7069 20-JUN-2001;
KYOWA HAKKO KOGYO CO., LTD. (JP)
FEATURES
    source
        1..309400
            /organism="Corynebacterium glutamicum"
            /db_xref="taxon:1718"
            /note="Seq 1 to long (3,309,400) split in 11, seq 7069
            3,000,001 3,309,400 309,400"
BASE COUNT 70133 a 86477 c 83115 g 69675 t
ORIGIN

```

```

alignment_scores:
    Quality: 498.50      Length: 480
    Ratio: 1.701        Gaps: 14
    Percent Similarity: 61.042      Percent Identity: 27.708
alignment_block:
US-09-911-667A-2 x AX127153 ..
Align seg 1/1 to: AX127153 from: 1 to: 309400
90 ThrGlyLeuAlaTrpMetAlaAspAlaMetGluMetIleLeuSerI1 106
:|||||:|||||:|||||:|||||:|||||:|||||:
179148 TCAGGTATCGGCTGGCATTTGGATGCCATGATCGGACTGATCTCTTT 179197
106 eLeuAlaProGlnLeuHisCysGluTrpArgLeuProSerTrpGlnValA 123
:|||||:|||||:|||||:|||||:|||||:|||||:
179198 CGTCATGGCTCGGTGGCCACTCATTTGGGCCTATCCCTACTGAAACTT 179247
123 laLeuLeuThrSerValValPheValGlyMetMetSerSerSerThrLeu 139
:|||||:|||||:|||||:|||||:|||||:|||||:
179248 CCCTGCTCGGATCCATCGGTTTCGTTGCATGGCATCGGCGCTTCGCTG 179297
140 TrpGlyAsnIleSerAspGlnTyrGlyArgLysThrGlyLeuLysIleSe 156
|||:|||||:|||||:|||||:|||||:|||||:
179298 GCGGTTTTCGTCGGCGGCAAGTTGGGCGCTCGA.....CAGTTT 179338
156 rValLeuTrpThrLeuTyrTyrGlyIle.....LeuSerAlaPheA 170
:|||||:|||||:|||||:|||||:|||||:|||||:
179339 TCGCTGTGCTTTTGTAGTTTATGGCGTTGCCACTGGCGCGTCGCGCTT 179388
170 laProValTyrSerTrpIleLeuValLeuArgGlyLeuValGlyPheGly 186
:|||||:|||||:|||||:|||||:|||||:|||||:
179389 CAGTCTCGCTGGCAATGTTGATGGCGCTGCGCTTCGCTGTTGGACTGGC 179438
187 lleGlyGly...ValProGlnSerValThrLeuTyrAlaGluPheLeuPr 202
:|||||:|||||:|||||:|||||:|||||:|||||:
179439 CTGGCGCTGAACTCCCGTTGCATCCACTCTGATTTCCGAGTTTCTCC 179488
202 oMetLysAlaArgAlaLysCysIleLeuLeuIleGluValPheTrpAlaI 219
:|||||:|||||:|||||:|||||:|||||:|||||:
179489 ACGAAAAAGTTCTGCGCGCATGCTGTTATCTGAGGCGGTTTTCGGCGC 179538
219 leGlyThrValPheGluValValLeuAlaValPheValMetProSerLeu 235
:|||||:|||||:|||||:|||||:|||||:|||||:
179539 TGGGCTGGATCATGCTGCAATCGTCGGAACCTTTCGTCGTCGAGATCC 179588
236 .....GlyTrpArgTrpLeuLeuIleLeuSerAlaValProLeuLeuLe 250
|||||:|||||:|||||:|||||:|||||:|||||:
179589 GATAATGTTGGCGTTGGCGTTAGCTCTTGGCTGTGCCCTGCAATTTA 179638
250 uPheAlaValLeuCysPheTrpLeuProGluSerAlaArgTyrAspValL 267
:|||||:|||||:|||||:|||||:|||||:|||||:
179639 CCGGTGTATGTCGCTCGGACTTCCAGAAATCAGTACGTTTCTCTGGAGA 179688
267 euSerGlyAsnGlnGluLysAla.....IleAlaThrLeuLysArg 280
:|||||:|||||:|||||:|||||:|||||:|||||:

```


Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-911-667A-2 x AA249686

Align seg 1/1 to: AA249686 from: 1 to: 2562

1 MetGluGluAspLeuPheGlnLeuArgGlnLeuProValValLysPheAr 17
 185 ATGGAGGAGGACTTATTCAGCTAGGACAGCTCCGGTGTGAAATTCG 234
 17 gArqThrGlyGluSerAlaArgSerGluAspThrAlaSerGlyGluH 34
 235 TCGCACAGCGAGTGCAGGTGCAAGGTGACAGGACGACCGCTTCAGGAGC 284
 34 isGluValGlnIleGluGlyValHisValGlyLeuGluAlaValGluLeu 50
 285 ATGAAGTCCAGATGAAGGGGTCCACGTGGGCTAGAGCTGTGAGAGCTG 334
 51 AspAspGlyAlaAlaValProLysGluPheAlaAsnProThrAspAspTh 67
 335 GATGATGGGACGCTGTGCCAAGGAGTTTGCCAAATCCACCGATGATAC 384
 67 rPheMetValGluAspAlaValGluAlaIleGlyPheGlyLysPheGlnT 84
 385 TTTTCATGGTGAAGATGCAGTGGAGCCATTCGCTTTGGAAATTCAGT 434
 84 rPlyLysLeuSerValLeuThrGlyLeuAlaTrpMetAlaAspAlaMetGlu 100
 435 GGAAGCTGTCTGCTCACTCGCTTGGCTGGATGGCTGATGCCATGGAG 484
 101 MetMetIleLeuSerIleLeuAlaProGlnLeuHisCysGluTrpArgLe 117
 485 ATGATGATCTCAAGCATCTCGGCACACACAGCTGCATTCGCGAGTGGAGGCT 534
 117 uProSerTrpGlnValAlaLeuLeuThrSerValValPheValGlyMetM 134
 535 CCCAAGCTGGCAGGTGCATTGCTGACCTCGGTGGTCTTTGTAGGCATGA 584
 134 etSerSerSerThrLeuTrpGlyAsnIleSerAspGlnTrpGlyArgLys 150
 585 TGTCCAGCTCCACGCTCTGGGGAAATATCTCAGACAGATACGGCAGGAAA 634
 151 ThrGlyLeuLysIleSerValLeuTrpThrLeuTrpTrpGlyIleLeuSe 167
 635 ACAGGCTGAACATCACGCTGCTGGACTGTACTATGSCATCCCTTAG 684
 167 rAlaPheAlaProValTyrSerTrpIleLeuValLeuArgGlyLeuValG 184
 685 TGCATTTGCGCCCGTGTATAGCTGGATCTCTGGTGTCTCCGGGGCTGGTGG 734
 184 lYPheGlyIleGlyValProGlnSerValThrLeuTrpAlaGluPhe 200
 735 GCTTCGGGATCGAGGAGTTCGCCAGTCGGTACGCTGTATGCCGAGTTC 784
 201 LeuProMetLysAlaArgAlaLysCysIleLeuLeuIleGluValPheTr 217
 785 CPTCCCATGAAGCCAGAGCTAAATGATTTGCTGATTGAGGTATTCTG 834
 217 pAlaIleGlyThrValPheGluValValLeuAlaValPheValMetProS 234
 835 GCCCATCGGACAGTGTTCGAGGTGCTGCTGGCTGTCTGCTGATGCCCA 884
 234 erLeuGlyTrpArgTrpLeuLeuIleLeuSerAlaValProLeuLeuLeu 250
 885 GCCTGGGCTGGCTGGCTGCTCATCTCTCAGCTGTCCGCTCCTCCTC 934
 251 PheAlaValLeuCysPheTrpLeuProGluSerAlaArgTyrAspValLe 267
 935 TTTGCGGCTGCTGCTTTCTGCTGCCCTGAAAGTCAAGGTATGATGCT 984
 267 uSerGlyAsnGlnGluLysAlaIleAlaThrLeuLysArgIleAlaThrG 284

985 GTCAGGGAAACGAGAAAGCAATCGCCACCTTAAAGAGGATAGCAACTG 1034
 284 luAsnGlyAlaProMetProLeuGlyLysLeuIleIleSerArgGlnGlu 300
 1035 AAAACGGAGCTCCCATCGCGCTGGGAAACTCATCATCTCCAGACAGGAA 1084
 301 AspArgGlyLysMetArgAspLeuPheThrProHisPheArgTrpThrTh 317
 1085 GACCGAGGCAAAATGAGGACCTTTTTCACACCCCATTTTAGTAGACAAAC 1134
 317 rLeuLeuLeuTrpPheIleTrpPheSerAsnAlaPheSerTrpTrpGlyL 334
 1135 TTTGCTGCTGCTGGTTATATGTTTTCATGCAATCTCTTACTACGGCT 1184
 334 euValLeuLeuThrThrGluLeuPheGlnAlaGlyAspValCysGlyIle 350
 1185 TAGTTCTACTCACACAGAACTCTTCCAGGAGGAGATGTCTGCGGCATC 1234
 351 SerSerArgLysLysAlaValGluAlaLysCysSerLeuAlaCysGluTy 367
 1235 TCCAGTCGGAAGAGCTGTAGAGGCAAAATCCACGCTGGCTGGCGACTA 1284
 367 rLeuSerGluGluAspTrpMetAspLeuLeuTrpThrThrLeuSerGluP 384
 1285 CCGTAGTGAGGAGGATTACATGGACTTGTGTGGACACCCCTCTCTGAGT 1334
 384 reProGlyValLeuValThrLeuTrpIleIleAspArgLeuGlyArgLys 400
 1335 TTTCCAGGTGCTCTGTGACTCTGTGGATTATGACCGCTGGGGCGCAAG 1384
 401 LysThrMetAlaLeuCysPheValIlePheSerPheCysSerLeuLeuLe 417
 1385 AAGACCATGGCCCTGTGCTTTGTCATCTTCTCTCTGACGCTCTCTGT 1434
 417 uPheIleCysValGlyArgAsnValLeuThrLeuLeuLeuPheIleAlaA 434
 1435 GTTATCTGCTGTGGAAGAAATGCTGCTCACTCTGTACTCTTTCATTCG 1484
 434 rGAlaPheIleSerGlyGlyPheGlnAlaIaIaValTrpValTrpProGlu 450
 1485 GAGCGTTTATTTCTGGAGCTTTCAAGCGCATATGTTTACACACCTGAG 1534
 451 ValTyrProThrAlaThrArgAlaLeuGlyLeuGlyThrCysSerGlyMe 467
 1535 GTCTACCCACGCAACGGGGGCTCTGGGCTGGGCACCTGACGGGCAT 1584
 467 lAlaArgValGlyAlaLeuIleThrProPheIleAlaGlnValMetLeuG 484
 1585 GGCAAGAGTGGTGTCTCATCATCTCCGTTTCATCGCCAGGTGATGCTGG 1634
 484 luSerSerValTyrLeuThrLeuAlaValTyrSerGlyCysCysLeuLeu 500
 1635 AATCCTCTGTGTACCTGACTCTGGCAGTTTACAGTGGCTGCTGCTCCTG 1684
 501 AlaAlaLeuAlaSerCysPheLeuProIleGluThrLysGlyGlyLe 517
 1685 GCTGCCCTGGCCTCCTGCTTTTTCCTCATTTGAGACCAAGGCGGAGACT 1734
 517 uGlnGluSerSerHisArgGluTrpGlyGlnGluMetValGlyArgGlyM 534
 1735 GCAGGAGTCCAGCCACCGGAGTGGGGCCAGGAGATGGTGGCGCCGAGGAA 1784
 534 ethisGlyAlaGlyValThrArgSerAsnSerGlySerGlnGlu 548
 1785 TGCACGCTGCAGCTGTACCAGGTGCAACTCTGGCTCTCAGGAA 1828

seq_name: /SIDS2/gcgdata/geneseq/geneseq.n/NA2000.DAT:AAZ49689

seq_documentation_block:

ID AAZ49689 standard; cDNA; 1906 BP.

XX

AC

XX

DT 07-APR-2000 (first entry)
XX Rat organic cation transporter-like protein (OCT1p) partial cDNA.
DE
XX Human; organic cation transporter-like protein; OCT1p; transporter;
KW transmembrane; norepinephrine; neuroprotective; neuroleptic; anticonvulsant;
KW antiparkinsonian; antidepressant; cellular process; cell proliferation;
KW screen; treatment; prevention; diagnosis; neurodegenerative disorder;
KW Alzheimer; Parkinson's; Huntington; ALS; amyotrophic lateral sclerosis;
KW CNS disorder; central nervous system; schizophrenia; depression;
KW behavioural; sleep disorder; eating disorder; rat; ss.
XX
OS Rattus sp.
XX
XX WO200000633-A1.
PN
XX 06-JAN-2000.
PD
XX 29-JUN-1999; 99WO-US14880.
PF
XX 30-JUN-1998; 98US-0107932.
PR
XX (MILL-) MILLENNIUM PHARM INC.
PA
XX Goodearl AJ, Glucksmann MA;
PI WPI; 2000-137069/12.
DR
XX New nucleic acid encoding human organic cation transporter-like
PT protein, used for prevention, treatment and diagnosis of e.g.
PT neurological, behavioural or sleep disorders
PT
XX Disclosure; Fig 4: 100pp; English.
XX
CC The present sequence is a partial cDNA encoding rat OCT1p (organic
CC cation transporter-like protein) and derived from rat frontal cortex
CC library. It was used to design primers to isolate OCT1p cDNA from human
CC foetal brain cDNA library. Human OCT1p is a member of the superfamily of
CC sugar and other transporter molecules that have 12 transmembrane
CC domains. It is highly expressed in brain tissue and has norepinephrine,
CC neuroprotective, neuroleptic, anticonvulsant, antiparkinsonian,
CC antidepressant activities. The OCT-like protein is used to
CC regulate a variety of cellular processes e.g. cell proliferation,
CC differentiation and survival, screen OCT1p modulators and detect mutation
CC in OCT1p gene. OCT1p modulators can be used to treat or prevent chronic
CC neurodegenerative disorders (e.g. Alzheimer's, Parkinson's,
CC Huntington's and ALS), CNS disorders (e.g. schizophrenia, panic,
CC depression), behavioural, sleep and eating disorders.
XX
SQ Sequence 1906 BP; 379 A; 547 C; 556 G; 424 T; 0 other;

alignment_scores:
Quality: 876.00 Length: 180
Ratio: 4.949 Gaps: 0
Percent Similarity: 98.333 Percent Identity: 93.889

alignment_block:
US-09-911-667A-2 x AA249689 ..

Align seg 1/1 to: AA249689 from: 1 to: 1906

369 SerGluGluAspTyrMetAspLeuLeuThrThrLeuSerGluPhePr 385
|||||
18 ACAGAGAGGATTACATGACCTGCTGTGGACCACTCTCTGAGTTCCC 67
|||||
385 oGlyValLeuValThrLeuTrpIleAspArgLeuGlyArgLysLysT 402
|||||
68 AGGTGCTCTGTGACTCTGTGGTCTATCGACCGCTGGCGCGAAGA 117
|||||
402 hrMetAlaLeuCysPheValIlePheSerPheCysSerLeuLeuPhe 418
|||||
118 CCATGGCTCTGTGTTCTGCTATCTTTTCCCTGACGCTCTCTGCTGTC 167

419 IleCysValGlyArgAsnValLeuThrLeuLeuLeuPheIleAlaArgAl 435
|||||
168 ATCTGCATTGCAAGAAATGTCTAACCTCTTACTGTTCATTGCAAGAGC 217
|||||
435 aPheIleSerGlyGlyPheGlnAlaAlaTyrValTyrThrProGluValT 452
|||||
218 GTTATTTCCTGGAGGCTTCCAAGCAGCTACGTTTACAGCGCTGAGGTGT 267
|||||
452 YrProThrAlaThrArgAlaLeuGlyLeuGlyThrCysSerGlyMetAla 468
|||||
268 ATCCAGCGGCGACGAGGCGCTGGCGCTGGCACCTGCAGCGCATGGCG 317
|||||
469 ArgValGlyAlaLeuLeuThrProPheIleAlaGlnValMetLeuGluSe 485
|||||
318 AGAGTGGCGCGCTCATCTACCTCCATTTCATAGCTCAGGTGATGCTGGAATC 367
|||||
485 rSerValTyrLeuThrLeuAlaValTyrSerGlyCysCysLeuLeuAlaA 502
|||||
368 TTCGCTGACTGACCTGCGCTGTACAGTGGCTGCTGCTGCTCTCTGCTG 417
|||||
502 laLeuAlaSerCysPheLeuProIleGluThrLysGlyGlyGlyLeuGln 518
|||||
418 CCTTGGCTCTCTGCTTCTGCCATCGAGACCAAGGCGGAGCAGCTGCGAG 467
|||||
519 GluSerSerHisArgGluTrpGlyGlnGluMetValGlyArgGlyMetHi 535
|||||
468 GAGTCCAGCCAGCGGAGTGGGCGCAGGAGATGGTTGCCGAGGAGACAA 517
|||||
535 sGlyAlaGlyValThrArgSerAsnSerGlySerGlnGlu 548
|||||
518 CAGCACAGGCGTCCCGAGTCTGCTGCTGCTCAGGAG 557

seq_name: /SDS2/gcdata/geneseq/geneseq/NA2000.DAT:AAA45312

seq_documentation_block:
ID AAA45312 standard; cDNA; 401 BP.

AC AAA45312;

XX 21-AUG-2000 (first entry)

XX Human secreted expressed sequence tag SEQ ID NO:1887.

XX Human; mouse; chicken; rat; secreted expressed sequence tag; sEST;
KW expressed sequence tag; EST; probe; chemotactic; proliferative;
KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW thrombolytic; antinflammatory; cytostatic; antibacterial; antifungal;
KW antiviral; antidiabetic; antiasthmatic; vulnary; antiparkinsonian;
KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW central nervous system disorder; Alzheimer's disease; stroke;
KW Parkinson's disease; Huntington's disease; coagulation disorder;
KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW tumour; infection; depression; psoriasis; ss.

XX Homo sapiens.

XX WO200021991-A1.

XX 20-APR-2000.

XX 15-OCT-1999; 99WO-US24206.

XX 15-OCT-1998; 98US-0104436.

XX (GEMY) GENETICS INC.

XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;

PI Merberg D, Treacy M, Bowman MR,

XX WPI: 2000-317938/27.
XX Isolated polynucleotides, and encoded proteins, comprising secreted
PT expressed sequence tags (sESTs), useful for treating various disorders
PT such as autoimmune, infectious, and central nervous system disorders -
XX
XX Claim 1: Page 679; 803pp; English.
XX AAA43426 to AAA45925 represent specifically claimed secreted expressed
CC sequence tags (sESTs), isolated from human, mouse, chicken and rat
CC tissue sources. The sESTs can have a range of activities depending on
CC the tissues they were isolated from. The activities include:
CC chemotactic; proliferative; immunomodulatory; haematopoietic;
CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
CC cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
CC antiasthmatic; vulnery; antitumor; osteoprotective; neuroprotective;
CC neurotropic; antiparkinsonian; antipsoriatic; cerebroprotective;
CC anticonvulsant; and antidepressant. The sESTs can be used for gene
CC therapy and in vaccines. The sESTs are useful as probes for the
CC identification and isolation of full-length cDNAs and genomic DNA
CC molecules which correspond to the sESTs. Proteins encoded by the sESTs
CC are useful in assays for determining biological activity and raising
CC antibodies. They may be useful for treatment of autoimmune disorders
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC osteoporosis, osteoarthritis, central nervous system disorders
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC disease), tumours, bacterial, fungal or viral infections, depression and
CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given
CC in the exemplification of the present invention.
XX
SQ Sequence 401 BP; 69 A; 119 C; 104 G; 109 T; 0 other;

alignment_scores:
Quality: 597.00 Length: 123
Ratio: 4.975 Gaps: 0
Percent Similarity: 97.561 Percent Identity: 96.748

alignment_block:

US-09-911-667A-2 x AAA45312 ..

Align seg 1/1 to: AAA45312 from: 1 to: 401

402 ThrMetAlaLeuCysPheValIlePheSerPheCysSerLeuLeuLeuPh 418
33 ACCATGGCCCTGTGCTTTGTTCATCTCTCTCTCTCTCTCTCTCTCTCT 82
418 IleCysValGlyArgAsnValLeuThrLeuLeuLeuPheIleAlaArgA 435
83 TATCTGTGTGGACAAATGTGCTCACCTCTGTACTCTTCTATTGCG 132
435 IapheIleSerGlyGlyPheGlnAlaAlaTyrValTyrThrProGluVal 451
133 CTTTATTCTCGAGGCTTTCAAGGGGATATGTTTACACACCTGAGTGC 182
452 TyrProThrAlaThrArgAlaLeuGlyLeuGlyThrCysSerGlyMetAl 468
183 TACCCACCGCAACGCGGCGCTCGGCTGGCACCTGCAGCGCATGGC 232
468 aArgValGlyAlaLeuIleThrProPheIleAlaGlnValMetLeuGluS 485
233 AAGAGTGGGTGCTCTCATCTACCTCCGTTTCATGCCAGGTGATGCTGGA 282
485 erSerValTyrLeuThrLeuAlaValTyrSerGlyCysCysLeuLeuAla 501
283 CTTCTGTACTGACTCTGCGAGTTTACAGTGGCTGCTGCTCTGCTGCT 332
502 AlaLeuAlaSerCysPheLeuProIleGluThrLysGlyGlyLeuGlu 518
333 GCCTTGGCTCTCTGCTTTTGGCCATTGAGACCAAGGCGGAGGACTGCA 382

518 nGluSerSerHisArgGlu 524
|||||:::
383 GGAGTCCCAAGCCACTCGAG 401

seq_name: /SDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH68534

seq_documentation_block:

ID AAH68534 standard; DNA; 309400 BP.

XX AAH68534;

XX 26-SEP-2001 (first entry)

XX C glutamicum coding sequence fragment SEQ ID NO: 7069.

XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.

XX Corynebacterium glutamicum.

XX EP1108790-A2.

XX 20-JUN-2001.

XX 18-DEC-2000; 2000EP-0127688.

XX 16-DEC-1999; 99JP-0377484.

XX 07-APR-2000; 2000JP-0159162.

XX 03-AUG-2000; 2000JP-0280988.

XX (KYOW) KY. I.A. HAKKU KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI: 2001-376931/40.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT

XX Disclosure: SEQ ID NO: 7069; 245pp + Sequence Listing: English.

XX The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.

XX Sequence 309400 BP; 70133 A; 86477 C; 83115 G; 69675 T; 0 other;

alignment_scores:

Quality: 498.50 Length: 480
Ratio: 1.701 Gaps: 14
Percent Similarity: 61.042 Percent Identity: 27.708

alignment_block:

US-09-911-667A-2 x AAH68534 ..

Align seg 1/1 to: AAH68534 from: 1 to: 309400

90 ThrGlyLeuAlaTyrMetAlaAspAlaMetGluMetIleLeuSer11 106

179148 TCAGGTATCGGCTGGGCATGATGCCATGGATGTCGAACTGATCTCTTT
 106 eLeuAlaProGlnLeuHisCysGluTrpArgLeuProSerTrpGlnValA 123
 179198 CGTCATGGCTGGTTGGCCACTCATTTGGGGCTATCCCTACTGAACTT 179247
 123 laLeuLeuThrSerValValPheValGlyMetMetSerSerThrLeu 139
 179248 CCCTGCTCGGATCCATCGTTTCGTTGGCATGGCCATCGGGCTTCGCTG
 140 TrpGlyAsnIleSerAspGlnTyrGlyArgLysThrGlyLeuLysIleSe 156
 179298 GCGGGTTGCTGGCGGCAAGTTGGGGCTCGA.....CAGATTTT 179338
 156 rValLeuTrpThrLeuTyrTyrGlyLe.....LeuSerAlaPheA 170
 179339 TCGGTTGCTTTGCTAGTTTATGGCTTGCCTCGCGCTCGGGCGCTT 179388
 170 laProValTyrSerTrpIleLeuValLeuArgGlyLeuValGlyPheGly 186
 179389 CAGTGTCTGCGTCAATGTGTATGGCGCTCGCGCTCGTCTGGACTTCGC 179438
 187 IleGlyGly...ValProGlnSerValThrLeuTyrAlaGluPheLeuPr 202
 179439 CTGGGGCTGAACCTCCCGTTGCATCCACTCTGATTCCGAGTTTCTCC 179488
 202 oMetLysAlaArgAlaLysCysIleLeuLeuIleLeuSerAlaValPheTrpAlaI 219
 179489 ACCAAAAGTTCGTGGCGCATGCTGTTATCTCTGGAGCGCTTTGGCGC 179538
 219 leGlyThrValPheGluValValLeuAlaValPheValMetProSerLeu 235
 179539 TGGGCTGGATCATGGCTGCAATCGTCGGAACCTTTGTCTGGCAGATGCC 179588
 236GlyTrpArgTrpLeuLeuIleLeuSerAlaValProLeuLeuLe 250
 179589 GATAATGGTTGGCTTGGCGCTAGTCTCTGTGCTGTGCCCTGCAATTA 179638
 250 uPheAlaValLeuCysPheTrpLeuProGluSerAlaArgTyrAspValL 267
 179639 CGCGGTGATGTCCGCTCGGACTTCCAGAATCAGTACTGTTCTCTCGAGA 179688
 267 euSerGlyAsnGlnGluLysAla.....IleAlaThrLeuLysArg 280
 179689 AGAAGCGCGGCACGACGAGCAGAACCATCGTCTTCTCTCGAAGAA 179738
 281 IleAlaThrGluAsnGlyAlaProMetProLeuGlyLysLeuIleIleSe 297
 179739 GCTCGCGCTCGCGAAGTAAAGCTGCCGATGCCACCGCTGCTGTTCA 179788
 297 rArgGlnGluAspArgGlyLysMetArgAspLeuPheThrProHisPheA 314
 179789 TGACAACGCTGCCGAGGGTTCCGTA...TCCATCTGGCTCAGCTGCTTGC 179835
 314 rqtTrpThrThrLeuLeuLeuTrpPheIleTrpPheSerAsnAlaPheSer 330
 179836 GCAAGCGCACCGCTCGCGCTGTGGATCGTGTGCTGCTGCAATCAACTTGC 179885
 331 TyrTyrGlyLeuValLeuLeuThrThrGluLeuPheGlnAlaGlyAspVa 347
 179886 TACTACGAGACCTTCATTTGGATTCCATCGCTGCTGCTGCTGCGACGGTT 179935
 347 lCysGlyIleSerSerArgLysLysAlaValGluAlaLysCysSerLeuA 364
 179936 CACCCTCGTGAAGTCT..... 179951
 364 lacCysGluTyrLeuSerGluGluAspTyrMetAspLeuLeuTrpThr 380
 179952TCCCAATTCATCTGATCATCACCC 179975
 381 LeuSerGluPheProGlyValLeuValThrLeuTrpIleIleAspArgLe 397

```

179976 TTGGCTCAGCTTCAGGCTATCCGGTTGCAGCGTGGTTGATTGAAAAGTG 180025
397 uclyArgLysLysThrMetAlaLeuCysPheValIlePheSerPheCysS 414
|||||:::|||||::: |||||::: |||||::: |||||::: |
180026 GGGCGGTGCGACACATTGGCC...ACGTTCCTGGTTGGTTCTGTATCT 180072
414 erLeuLeuLeuPheIleCysValGlyArgAsnValIleThrLeuLeuLeu 430
|| |||||::: |||||::: |||||::: |||||::: |||||::: |
180073 CTGCAGCGCTCTAC.....GGCTGGCAAAATGTGGAGTGGCAGATG 180113
431 PheIleAlaArgAlaPheIleSer.....GlyGlyPheGlnAl 443
|||||:::|||||::: |||||::: |||||::: |||||::: |||||::: |
180114 CTGGTCGCAGGTGTGTTACTATCTCTTCAACCTGGCGCATGGGCGC 180163
443 alaTyrrValTyrrProGluValTyrrProThrAlaThrArgAlaLeuG 460
|| |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |
180164 ACTGATGCCAATTGGCGCGAGCTTTATCCCACATAATGTCCTGGGAAC 180213
460 lyLeuGlyThrCysSerGlyMetAlaArgValGlyAlaLeuIleThrPro 476
|| |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |
180214 GAACGGTGCTCGCGGGGATTGGGGCGATTGCTTCCATCATCGCTCCG 180263
477 PheIleAlaGlnValMetLeuGluSerSerValTyrrLeuThrLeu...Al 492
|||||:::|||||::: |||||::: |||||::: |||||::: |||||::: |
180264 CTCATTGTTCCCGCACTGATTGTTTTGGTGGACCAATTCGTTTGTTCG 180313
492 avallTyrrSerGlyCysLeuLeuAlaAlaLeuAlaSerCysPheLeuP 509
|||||:::|||||::: |||||::: |||||::: |||||::: |||||::: |
180314 TCTCTTTGGCACCGGCTTTGGCGGATTCGACGAGTTGCGGCTTTACGCTGC 180363
509 rolleGluThrLysGlyGlyLeuGlnGlu..... 519
|| |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |
180364 CT...GAGCAGAAGGTAAGTCTCTCGCTGATTAGTGAGATCCTCCACC 180410
520 .....SerSerHisArgGluThrGlyGlnGluMetValGlyArgGly.W 534
||| :::::|||||::: |||||::: |||||::: |||||::: |||||::: |
180411 AGTTTTTCACCTGCCCAACGATGGTGC GTTCTGCGCCTGGCGGAAGTCC 180460
534 exHisGlyAlaGlyValThrArgSerAsnSerGlySer 546
:::::|||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |
180461 TCCAATTCGGGGGAGGTTCGGGATAACCGTGGGCACC 180498

seq_name: /SIDSt/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH68260
seq_documentation_block:
seq_AAH68260 standard; DNA; 1344 BP.
AC AAH68260;
DDT (first entry)
26-SEP-2001 (first entry)
C glutamicum coding sequence fragment SEQ ID NO: 3295.
Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
organic acid synthesis; ds.
Corynebacterium glutamicum.
EP1108790-A2.
20-JUN-2001.
18-DEC-2000; 2000EP-0127688.
16-DEC-1999; 99JP-0377484.
07-APR-2000; 2000JP-0159162.
03-AUG-2000; 2000JP-0280988.
(KYOW ) KYOWA HAKKO KOGYO KK.
Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi
Tateishi N, Senoh A, Ikeda M, Ozaki A;

```

seq name: /SIDS2/qcdata/geneseq/geneseq/NA2001.DAT:AAF67999


```

|||||.....|
316 GATGATGGGGCAKCTGTGCCCAAGAGTTTGCCTCAATCCACCGATGATAC 365
67 rPheMetValGluAspAlaValGluAlaIleGlyPheGlyLysPheGlnT 84
|||||.....|
366 TTTCATGTTGGAAGATGAGTGGAGCCATTTGGCTTTGGAAAAATTCAGT 415
84 rPlysLeuSerValLeuThrGlyLeuAlaTrpMetAlaAspAlaMetGlu 100
|||||.....|
416 GGAAGCTGTCTTCTCAATGCTTGGCTTGATGCT.SATGCCATGAG 464
101 MetMetIle 103
|||||
465 ATGATGATC 473

```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ44679

seq_documentation_block:

AAZ44679 standard; cDNA to mRNA; 1901 BP.

```

XX AC AA244679;
XX DT 14-APR-2000 (first entry)
XX DE Rat liver anion transporter protein OAT2 encoding cDNA.
XX KW OAT2; anion transporter; liver; rat; dicarboxylic acid; prostaglandin;
XX KW non-steroid anti-inflammatory agent; anti-tumor; ss.
XX OS Rattus sp.

```

```

XX FH Key Location/Qualifiers
XX CDS 51..1659
XX FT /*tag= a
XX FT /product= "OAT2"

```

PN JP11346779-A.

XX 21-DEC-1999.

XX PF 03-JUN-1998; 98JP-0169174.

XX PR 03-JUN-1998; 98JP-0169174.

XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.

XX WPI; 2000-109494/10.

DR P-PSDB; AAY51249.

XX A liver-specific org. anion transporter - can transport various substances including dicarboxylic acids

XX Claim 4; Page 6-7; 13pp; Japanese.

XX This invention describes a novel rat liver-specific org. anion transporter OAT2. The liver-specific org. anion transporter OAT2 can transport various substances including dicarboxylic acids, prostaglandins, non-steroid anti-inflammatory agents and anti-tumor agents. This sequence encodes the rat liver anion transporter OAT2 protein described in the method of the invention.

XX Sequence 1901 BP; 346 A; 537 C; 571 G; 447 T; 0 other;

alignment_scores:

```

Quality: 446.50 Length: 552
Ratio: 1.493 Gaps: 18
Percent Similarity: 54.167 Percent Identity: 27.899

```

alignment_block:

US-09-911-667A-2 x AAZ44679 ..

Align seg 1/1 to: AAZ44679 from: 1 to: 1901

```

34 HisGluValGlnIleGlyValHisValGlyLeuGluAlaValGluLe 50
|||||.....|
189 CACATTGTGCCCTGCCGTGGCCCTGCCCACTCAGTCACCGAGCTT 238
50 uAspAspGlyAlaAlaValProLysGluPheAlaAsnProThrAspSPT 67
|||||.....|
239 ATGCTGGAAGCCCATCTACCCCGGAG.....ACTGACGGCA 276
67 hrPheMet.....Val 70
|||||
277 GCTTTAGCTCTGCCTCGATTTGCCCTATCCCGACAGCTGTCCCAATGTC 326
71 GluAspAlaValGluAlaIleGlyPheGlyLysPheGlnTrpLys....Le 86
|||||.....|
327 ACTTTGGGGACAGAGGTCTCCAACTCTGGGGAGGCTGAGGCTGAGCCCT 376
86 uSerValLeuThrGlyLeuAlaTrpMetAlaAspAlaMetGluMetMet 103
|||||.....|
377 CACGTCGCCCTGCTCAGGGCTGGAGTACGACCGCTCAGAA..... 419
103 LeuSerIleLeuAlaProGlnLeuHisCysGluTrpArgLeuProSer 119
|||||.....|
420 .....TTCTCTCCACCATTTGCACTGAGTGGGATCTTGTGTGT 458
120 TrpGlnValAlaLeu.....LeuThrSerVal.....ValPheValGlyMe 133
|||||.....|
459 CAGCAGAGAGACTCAACAAATTTACGTCCACCTGCTTCTTCATTTGGTGT 508
133 tMetSerSerThrLeuTrpGlyAsnIleSerAspGlnTrpGlyArgL 150
|||||.....|
509 GCTGGTGG:AGCCGTGGGTATGATACTTGTCTGACAGGTTTGCAGGC 558
150 ysThrGlyLeuLysIleSerValLeuTrpThrLeuTrpGlyIleLeu 166
|||||.....|
559 GCCGCTTCTGCTGCTGCTACGTGAGTCCCTGCTGCTGGCTGTGATG 608
167 SerAlaPheAlaProValTyrSerTrpIleLeuValLeuArgGlyLeuVa 183
|||||.....|
609 TCTGCAGGCTCCATCACTATCATCATGTTCTGTAGTCACCCGTACACTCAC 658
183 tGlyPheGlyIleGlyValProGlnSerVal...ThrLeuTyrAlaG 199
|||||.....|
659 CGGCTCAGCCCTGGCTGCTTCCACCATCATTTGCTGCCACTGGAGTTGG 708
199 LuPheLeuProMetLysAlaArgAlaLysCysIleLeuLeuIleGluVal 215
|||||.....|
709 AGTGGCTGGATGTGGAGCAGCCGCTGTGGCCGGGTCTATCAGCACCGTTC 758
216 PheTrpAlaIleGlyThrValPheGluValValLeuAlaValPheValMe 232
|||||.....|
759 TTCTGCTG.....GGAGGCGTGGCTGCTGGTGGCAGCTGGTGGCTACCTGAT 805
232 tProSerLeuGlyTrpArgTrpLeuLeuIleLeuSerAlaValProLeuL 249
|||||.....|
806 CGGAGAGC.....TGCGCTGGCTTCTGCTGGCTGCCACCTGCCGTGGT 849
249 euLeuPheAlaValLeuCysPheTrpLeuProGluSerAlaArgTyrAsp 265
|||||.....|
850 TCCCAGGCATCATCAGCATCTGGTGGTTCCTGAGTCTGCACGGTGGCTT 899
266 ValLeuSerGlyAsnGlnGluLysAlaIleAlaThrLeuLysArgIleAl 282
|||||.....|
900 CTAACCCAGGGTGTGTGGAGGAGGCAAAAAAATACTTGTGAGCTGTGC 949
282 aThrGluAsnGlyAlaProMetProLeuGlyLysLeu..... 294
|||||.....|
950 CAAGCTCAATGGCGCGCGCTGGGTGGGTGAGGCGACCTTGAGCCAGGAGCCC 999
295 .....IleIleSerArgGlnGluAspArgGlyLysMetArg 306
|||||.....|
1000 TGAACAACGTGTGCTCACCATGGAAAGGGCGTTGCAAAAGACCCCTCATACTTA 1049

```

```

307 AspLeuPhe...ThrProHisPheArgTrpThrThrLeuLeuLeuTrpPh 322
||||| ||| : : : : : ||| : : : : :
1050 GACCTGTTCCGAACATCTACGCTCGACATATCTACCTGCTGCATGAT 1099
322 eiletrpPheSerAsnAlaPheSerTyrTyrGlyLeuValLeuLeuThr 339
: : : : : ||| : : : : : ||| : : : : :
1100 GGTGTGTTTGGAGTGAACCTTTCTACTACGCGCTGACTCTG..... 1142
339 hrGluLeuPheGlnAlaGlyAspValCysGlyIleSerSerArgLysLys 355
||||| |||
1143 .....GACGTGCTGGG..... 1154
356 AlaValGluAlaLysCysSerLeuAlaCysGluTyrLeuSerGluGluAs 372
||||| : : : : :
1155 .....CTGGGCTGAACGT 1168
372 pTyrMetAspLeuLeuTrpThrThrLeuSerGluPheProGlyValLeuV 389
||||| ||| : : : : : ||| : : : : :
1169 GTACCAGACACAGCTGCTGTTTGGGGCTCTGGAGCTCCCTCCAAATTA 1218
389 alThrLeuTrpIleIleAspArgLeuGlyArgLysLysThrMetAla... 404
: : : : : ||| : : : : : ||| : : : : :
1219 TGTCTACTTCTGTGGCGGCTCTGGAGCGCTCTCAGGAGCTGG 1268
405 ...LeuCysPheValIlePheSerPheCysSerLeuLeuLeuPheIleCy 420
||||| : : : : : ||| : : : : :
1269 ATGCTGCTGGCGCTGCTCTGACCTTGGCACCGCTGCTGATCTCTT 1318
420 sValGlyArgAsnValLeuThrLeuLeuPheIleAlaArgAlaPheI 437
: : : : : ||| : : : : : ||| : : : : :
1319 GGAGCTAAGTATGATGATGATGCTGCTGTGGTGGTGGGAAGCTTTT 1368
437 leSerGlyGlyGlnAlaAlaTyrValTyrThrProGluValTyrPro 453
: : : : : ||| : : : : : ||| : : : : :
1369 CTGAAGCTCTTTTACTACGGCTTACCTGTTCCAGTCCGAGTTGACCT 1418
454 ThrAlaThrArgAlaLeuGlyLeuGlyThrCysSerGlyMetAlaArgVa 470
||||| ||| : : : : : ||| : : : : :
1419 ACTGTGCTGACAGACAGAGATTGGGACTTACTGCACATCGGCGAGCT 1468
470 lGlyAlaLeuIleThrProPheIleAlaGlnValMetLeuGluSerSerV 487
||||| : : : : : ||| : : : : :
1469 AGGGGCTCTGTCGCCCTGCTGCGGCC.....TTGCTGGATGGAG 1509
487 alTyrLeuThrLeu.....AlaValTyrSerGlyCysCysLeuAla 501
||||| ||| : : : : : ||| : : : : :
1510 TGTGGCTGTGTGCTGCCAAAGTTGCTTACGGGGGATTGCCCTGTGGCT 1559
502 AlaLeuAlaSerCysPheLeuProIleGluThrLysGlyGlyGluG1 518
||||| : : : : : ||| : : : : :
1560 GCTGCACCTGCCTGCTGCT...GAGACGAGAGGACAGCTGCC 1606
518 nGluSer..... 520
|||||
1607 AGAGACCATCCAGGATGTGGAGGAAGAGTACCCAGGAGGAGATGTGT 1656
521 .....SerHisArgGluTrpGly 526
||||| : : : : : ||| : : : : :
1657 AGGTCCGGGACTGAGTTGCTAGCAGACAGTCTCCACAGGAGCTGGGCA 1706
527 GlnGlu 528
|||||
1707 CACAAG 1712

seq_name: /SID52/cgdata/geneseq/geneseq/NAL1996.DAT: AAT08702

seq_documentation_block:
ID AAT08702 standard; DNA; 1882 BP.
XX
AC AAT08702;
XX
DT 10-SEP-1996 (first entry)
XX
```

449 ProGluValTyrProThrAlaThrArgAlaLeuGlyLeuGlyThrCysSe 465
 1400 GTGAGCTGTACCTACATTCATCATCAGGAATCTTGGATGATGATGCTC 1449
 465 rGlyMetAlaArgValGlyAlaLeuLeuThrProPheLeuAlaGlnValM 482
 1450 TGCCTGTGTGACCTGGGTGGGATCTTACCCCTTCATGTTGTTTCAGGC 1499
 482 etLeuGluSerSerValTyrLeuThrLeuAlaValTyrSerGlyCysCys 498
 1500 TGATGGAAGTTTGGCAAGCCCTGCCCTCATTTTGTGGGTGTTTGGGC 1549
 499 LeuLeuAlaAlaLeuAlaSerCysPheLeuProIleGluThrLysGlyG 515
 1550 CTGACTGTGGGCCCATGACTCTCTCTCCCA...GAGACCAAGGTGT 1596
 515 yGlyLeuGlnGluSer.....Ser.HisArgGluTrpGly.GlnGluMe 529
 1597 GGCTTTGCCTGAGACTATTGAAGAAGCAGAGAACCCTGGGGAGGAGAAAT 1646
 529 tValGlyArgGlyMeuHis 535
 1647 CAAAGGCCAAAGAAACAC 1665
 seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:AAV49559
 seq_documentation_block:
 ID AAV49559 standard; cDNA to mRNA; 1662 BP.
 XX
 AC AAV49559;
 XX
 DT 21-OCT-1998 (first entry)
 XX
 DE Human liver cell clone HP01293 cDNA #2.
 XX
 KW Transmembrane domain; human; nutrition; cytokine; cell proliferation;
 KW differentiation; immune system; stimulator; suppressor; regulator;
 KW hematopoiesis; activin; inhibitor; chemotactic; chemokinetic; receptor;
 KW haemostatic; thrombolytic; ligand; anti-inflammatory; tumour; ds.
 XX
 OS Homo sapiens.
 XX
 PN W09821328-A2.
 XX
 PD 22-MAY-1998.
 XX
 PF 07-NOV-1997; 97WO-JP04056.
 XX
 PR 13-NOV-1996; 96Jp-0301429.
 XX
 PA (PROT-) PROTEGENE INC.
 PA (SAGA) SAGAMI CHEM RES CENTRE.
 XX
 PI Kato S, Kobayashi M, Sekine S, Yamaguchi T;
 XX
 DR WPI; 1998-297932/26.
 DR P-PSDB; AAW64538.
 XX
 PT Human protein having transmembrane domain - useful for, e.g.
 PT research and nutrition
 XX
 PS Claim 3; Page 120-121; 205pp; English.
 XX
 CC AAV49550-V49599 are cDNA sequences which encode human proteins containing
 CC a transmembrane domain. These proteins can be used for, e.g. research
 CC and nutrition, and may have cytokine and cell
 CC proliferation/differentiation, immune stimulating/suppressing,
 CC hematopoiesis regulating, tissue growth, activin/inhibin,
 CC chemotactic/chemokinetic, haemostatic and thrombolytic, receptor/ligand,
 CC anti-inflammatory or tumour inhibition activity.
 XX
 SQ Sequence 1662 BP; 312 A; 486 C; 466 G; 398 T; 0 other;

169 PheAlaProValTyrSerTrpIleLeuValLeuArgGlyLeuValGlyPhe 185
 623 GTGGCCCAAGACTATACATCCATGTTGCTTTCGGCTGCTGCAGGCAT 672
 185 e...GlyIleGlyValProGlnSerValThrLeuTyrAlaGluPheL 201
 673 GGTGAGCAAGGCGAGCTGGTGTCCGGCTATACCTTGCATCAGAGTTTG 722
 201 euProMetLysAlaArgAlaLysCysIleLeuLeuIleGluValPheTrp 217
 723 TCGGCTCTGCTACAGAGAGAACAGCGCCATTTTACACAGATGGCCTTC 772
 218 AlaIleGlyThrValPheGluValValLeuAlaValPheValMetProSe 234
 773 ACAGTGGGTAGTGGGCTTCCGGGTGGCC...TATGCCATTCAGA 819
 234 rLeuGlyTrpArgTrpLeuLeuIleLeuSerAlaValProLeuLeuLeuP 251
 820 C.....TGGCGTGGCTCCAGCTAGCTGTGCTCCCTCCCTACCTTCCTCT 863
 251 heAlaValLeuCysPheTrpLeuProGluSerAlaArgTyrAspValLeu 267
 864 TCCTGCTGTTACTGCTTGTCCAGAAATCCGCCCGGTGCTGTCTGTC 913
 268 SerGlyAsnGlnGluLysAlaIleAlaThrLeuLysArgIleAlaThrG 284
 914 CAGAGAGAACCCAGCGAGCTGTCCAGGATAATGGAGCAAAATGGCACAGA 963
 284 uAsnGlyAlaProMetProLeuGlyLysLeuIleIleSerArgGlnGlu 301
 964 GAACGGGAAGTGCTCTCTGCTGACCTGAAGATGCTGCTTGGAGAG 1013
 301 spArgGlyLysMetArg.....AspLeuPhe.....ThrPro 311
 1014 ATGCCTCGAAGAGAGAGTCTTGGTGGCGACCTTCCGACATCC 1063
 312 HisPheArgTrpThrLeuLeuLeuTrpPheIleTrpPheSerAsnAl 328
 1064 ACCCTGAGGAACACACGCTCATCTGATGATATCTATGTTCTCTGTGC 1113
 328 aPheSerTyrTyrGlyLeuValLeuThrThrGluLeuPheGlnAlaG 345
 1114 TGTGCTGTACAGGCTCTCATCATG..... 1138
 345 lYaspValCysGlyIleSerSerArgLysLysAlaValGluAlaLysCys 361
 1139CACGTGGAGCCACAGGGGCCAACCTC 1165
 362 SerLeuAlaCysGluTyrLeuSerGluGluAspTyrMetAspLeuTr 378
 1166TACCTGGACTCTTTTA 1182
 378 pThrThrLeuSerGluPheProGlyValLeuValThrLeuTrpIleIleA 395
 1183 TTCTTCTCTGGTGAATTCCTCGCGGCTTCATCATCTCTGGTCACCATG 1232
 395 spArgLeuGlyArgLysLysThrMetAlaLysCys...PheValIlePhe 410
 1233 ACCGATTGGCGCATCTACCAATAGCGGCTCGAATCTGTGTGACGGG 1282
 411 SerPheCysSerLeuLeuLeuPheIle..... 419
 1283 GCAGCTGCTCTCATGATCTTATCCCGCATGAGCTGACTGTTGAA 1332
 420CysValGlyArgAsnValLeuThrLeuLeuLeuPheI 432
 1333 CGTTACCTCGCTGTCTTGGCGGTATGGGGGCCACCATTTGGCTG.... 1378
 432 leAlaArgAlaPheIleSerGlyPheGlnAlaAlaTyrValTyrThr 448
 1379CAGATGCTGCTGCTGTGAAC 1399

alignment_scores: Quality: 424.00 Length: 422
 Ratio: 1.767 Gaps: 11
Percent Similarity: 56.872 Percent Identity: 29.858

alignment_block:
US-09-911-667A-2 x AAV49559 ..

Align seg 1/1 to: AAV49559 from: 1 to: 1662

```
119 SerTrpGlnValAlaLeuLeuThrSerValValPheValGlyMetMetse 135
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
433 TCCTGGAAGCTGACCTCTTTTCAGTCTCTTTGTAATGCGGGCTCTCTT 482
135 rSerSerThrLeuTrpGlyAsnIleSerAspGlnTrpGlyArgLysThrG 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
483 TGCTCTCTCGGTGGTGGCTACTTTGCAGACAGAGTTTGGCCGTAAGCTGT 532
152 lLeuLysIleSerValLeuTrpThrLeuTrpGlyIleLeuSerAla 168
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
533 GTCTCTGGGAAGCTGTGCTGGTCAACGGCGTGTGCGGCTGCTCATGGCC 582
169 PheAlaProValTrpSerTrpIleLeuValLeuArgGlyLeuValGlyPh 185
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
583 TTCTGCCCAACTACATGTCCATGTGCTCTTCCGCCCTGCTGCAGGGCT 632
185 e...GlyIleGlyValProGlnSerValThrLeuTrpGlyIleLeuPhe 201
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
633 GGTACAGCAAGGCACTGTGCTGGCTACACCTTAATCAGAAATTG 682
201 euProMetLysAlaArgAlaLysCysIleLeuLeuIleGluValPheTrp 217
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
683 TTGGCTCGGGCTCCAGAAAGCGGTGGCGATCATGTACAGATGGCTTC 732
218 AlaIleGlyThrValPheGluValValLeuAlaValPheValMetProse 234
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
733 ACGTGGGGCTGGTGGCGCTTACCGGGCTGGCC...TACGCCCTGCCTCA 779
234 rLeuGlyTrpArgTrpLeuLeuLeuSerAlaValProLeuLeuLeuP 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
780 C.....TGGCGTGGCTGCAGCTGGCAGTCTCCCTGCGCCACTCTCTCT 823
251 heAlaValLeuCysPheTrpLeuProGlnSerAlaArgTrpAspValLeu 267
|| : : : : : : : : : : : : : : : : : : : : : : : : : : :
824 TCCTGTCTACTACTGTGTGTCGCGGAGTCCCTCGGTGGCTGTATCA 873
268 SerGlyAsnGlnGlyAlaIleAlaThrLeuLysArgIleAlaThrG 284
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
874 CAAAAAAGAAACACTGAAGCAATAAAGATAATGGACCATCGCTCAAAA 923
284 uAsnGlyAlaProMetProLeuGlyLysLeuIleIleSerArgGlnGlu 301
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
924 GAATGGGAAGTGCCTCTCTGCTGATTTAAAGATGCTTCCCTCGAAGAGG 973
301 sp.....ArgGlyLysMetArgAspLeuPhe...ThrPro 311
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
974 ATGTCACCGAAAAAGCTGAGCCCTTCATTTCAGACCTGTTCCGACGCGG 1023
312 HisPheArgTrpThrThrLeuLeuLeuTrpPheIleTrpPheSerAsnAl 328
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1024 CGCTGTAGGAAGCGCACCTTCATCCTGTGATGACCTGTGTTCCGGACTC 1073
328 aPheSerTrpTrpGlyLeuValLeuLeuThrThrGluLeuPheGlnAlaG 345
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1074 TGTGCTCTATCAGGGGCTCATCTG.....CACATGG 1105
345 lAspValCysGlyIleSerSerArgLysLysAlaValGluAlaLysCys 361
|| : : : : : : : : : : : : : : : : : : : : : : : : : : :
1106 GCGCCACCAAGCGGG..... 1119
362 SerLeuAlaCysGlnTrpLeuSerGluGluAspTrpMetAspLeuLeuTr 378
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
1120 .....AACCTCTACCTGGATTTCCTTTA 1142
378 pThrThrLeuSerGluPheProGlyValLeuValThrLeuTrpIleIleA 395
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1143 CTCCTGCTGTGGTGAATCCCGGGCCCTTCATAGCCCTCATCACCATTG 1192
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
395 spArgLeuGlyArgLysLysThrMetAlaLeuCysPheValIlePhe... 410
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1193 ACCGGTGGCGGCATCTACCCCATGGCGTGTCAAATTTGTTGGCGGGG 1242
411 SerPheCysSerLeuLeuLeuPheIle..... 419
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1243 GCAGCTGCTGCTCATGATTATCTACCTGCACCTGCACCTGGTTAA 1292
420 .....CysValGlyArgAsnValLeuThrIleLeuLeuPheI 432
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1293 CATCATATCATGTGTGTTGGCGGAATGGAAATCACCATTGCAATA... 1338
432 leAlaArgAlaPheIleSerGlyPheGlnAlaAlaValTrpThr 448
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1339 .....CAAATGATCTGCCTGGTGAAT 1359
449 ProGluValTrpProThrAlaThrArgAlaLeuGlyLeuGlyThrCysSe 465
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1360 GCTGAGCTGTACCCACATTCGTCAGAACTCGGAGTATGTTGTTTC 1409
465 rGlyMetAlaArgValGlyAlaLeuIleThrPropheIleAlaGlnValM 482
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1410 CTCCTGTGTGACATAGTGGGATAATCACCCCTTCATAGTCTTCAGGC 1459
482 eLeuGluSerSerValTrpLeuThrLeuAlaValTrpSerGlyCysCys 498
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1460 TGAGGAGGTCTGGCAAGCTTGCCTCATTTTGTTCGGGTGGGC 1509
499 LeuLeuAlaAlaLeuAlaSerCysPheLeuProIleGluThrLysGly 515
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1510 CTGCTGCGCGGAGTGACGTACTTCTTCCA...GAGACCAAGGGGT 1556
515 yGlyLeuGlnGluSer 520
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1557 CGCTTTGCCAGAGACC 1572
```

seq_name: /SIDS2/gcdata/geneseq/geneseqn/NA1998.DAT:AAV49558

seq_documentation_block:

ID AAV49558 standard; cDNA to mRNA; 1888 BP.

XX AC AAV49558;

XX DT 21-OCT-1998 (first entry)

XX DE Human liver cell clone HP01293 cDNA #1.

XX KW Transmembrane domain; human; nutrition; cytokine; cell proliferation;
differentiation; immune system; stimulator; suppressor; regulator;
hematopoiesis; activin; inhibitor; chemotactic; chemokinetic; receptor;
hemostatic; thrombolytic; ligand; anti-inflammatory; tumour; ds.

OS Homo sapiens.

XX FH Key Location/Qualifiers

FT CDS 90..1745

FT FT /*tag= a

FT FT /product= "transmembrane domain containing protein"

PN WO9821328-A2.

XX PD 22-MAY-1998.

XX PF 07-NOV-1997; 97WO-JP04056.

XX PR 13-NOV-1996; 96JP-0301429.

Tue Mar 12 15:54:26 2002

PA (PROT-) PROTEGENE INC.
 PA (SAGA) SAGAMI CHEM RES CENTRE.
 XX Kato S, Kobayashi M, Sekine S, Yamaguchi T;
 PI WPI; 1998-297932/26.
 DR P-PSDB; AAW64538.
 XX Human protein having transmembrane domain - useful for, e.g.
 PT research and nutrition
 XX Claim 4; Page 141-143; 205pp; English.
 XX AAV49550-V49599 are cDNA sequences which encode human proteins containing
 CC a transmembrane domain. These proteins can be used for, e.g. research
 CC and nutrition, and may have cytokine and cell
 CC proliferation/differentiation, immune stimulating/suppressing,
 CC haematopoiesis regulating, tissue growth, activin/inhibin,
 CC chemotactic/chemokinetic, haemostatic and thrombolytic, receptor/ligand,
 CC anti-inflammatory or tumour inhibition activity.
 XX
 SQ Sequence 1888 BP; 365 A; 545 C; 520 G; 458 T; 0 other;

alignment_scores:
 Quality: 424.00 Length: 422
 Ratio: 1.767 Gaps: 11
 Percent Similarity: 56.872 Percent Identity: 29.858

alignment_block:
 US-09-911-667A-2 x AAV49558 ..

Align seg 1/1 to: AAV49558 from: 1 to: 1888

```

119 SerTrpGlnValAlaLeuLeuThrSerValValPheValGlyMetMetSe 135
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
522 TCCTGGAAGCTGGACCTCTTTCAGTCTCTTGAATGGCGGCTCTTCTT 571

135 rSerSerThrLeuTrpGlyAsnIleSerAspGlnTyrGlyArgLysThrG 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
572 TGGCTCTCTCGGTGTGGTCTACTTTCAGACAGGTTTGGCCGTAAAGCTGT 621

152 lLeuLysIleSerValLeuTrpThrLeuTyrTyrGlyIleLeuSerAla 168
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
622 GTCTCTGGGAACGTGCTGGTTCACGGCGGTCTCGGGGTGCTCATGGCC 671

169 PheAlaProValTyrSerTrpIleLeuValLeuArgGlyLeuValGlyPh 185
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
672 TTCTCCGCCCACTACATGTCCTGCTGCTCTCCGCTGCTGCAGGGCT 721

185 e...GlyIleGlyGlyValProGlnSerValThrLeuTyrAlaGluPheL 201
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
722 GGTACCAAGGGTACCTGGATGGCTGGCTACACCTAATCACAGAAATTG 771

201 euProMetLysAlaArgAlaLysCysIleLeuLeuIleGluValPheTrp 217
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
772 TTGGCTCGGGCTCCAGAAAGCGGTGGCGATCATGTACAGATGGCCTTC 821

218 AlaIleGlyThrValPheGluValValLeuAlaValPheValMetProSe 234
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
822 ACGGTGGGGCTGGTGGCGCTTACCGGGCTGGCC...TAGCCCTCGCCTCA 868

234 rLeuGlyTyrArgTrpIleLeuIleLeuSerAlaValProLeuLeuLeuP 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
869 C.....TGGCGCTGGCTGCAGCTGGCAGTCTCCCTGCCACCTTCTCT 912

251 heAlaValLeuCysPheTrpLeuProGluSerAlaArgTyrAspValLeu 267
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
913 TCCTGTCTACTACTGTGTGGCGGAGTCCCTCGGTGGCTGTATCA 962

268 SerGlyAsnGlnGluLysAlaIleAlaThrLeuLysArgIleAlaThrGl 284
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
963 CAAAAAAGAAACACTGAGCAATAAAGATAATGGACCACATCGCTCAAAA 1012

```

```

284 uAsnGlyAlaProMetProLeuGlyLysLeuIleIleSerArgGlnGluA 301
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1013 GAATGGGAAGTTGCTCTCTCTGATTTAAAGATGCTTTCCCTCGAAGAGG 1062

301 sp.....ArgGlyLysMetArgAspLeuPhe...ThrPro 311
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1063 ATGTCACCGAAAGAGCTGAGCCCTTCATTTGCAGACCTGTTCCGCACGCG 1112

312 HisPheArgTrpThrLeuLeuLeuLeuTrpPheIleTrpPheSerAsnAl 328
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1113 CGCTGAGGAAGCGCACCTTCATCTGATGTACCTGTGTGTTCACGGACTC 1162

328 aPheSerTyrTyrGlyLeuValLeuLeuThrThrGluLeuPheGlnAlaG 345
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1163 TGTGCTCTATACAGGGGCTCATCCTG.....CACATGG 1194

345 lYAspValCysGlyIleSerSerArgLysLysAlaValGluAlaLysCys 361
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1195 GCGCCACACGCGG..... 1208

362 SerLeuAlaCysGluTyrLeuSerGluGluAspTyrMetAspLeuLeuTr 378
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1209 .....AACCTCTACCTGGATTTCCCTTTA 1231

378 pThrThrLeuSerGluPheProGlyValLeuVal101hrLeuTrpIleIleA 395
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1232 CTCGCTCTGTGCAAAATCCCGGGGCTTCATAGCCCTCATCACCATTG 1281

395 sPArgLeuGlyArgLysLysThrMetAlaLeuCysPheValIlePhe... 410
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1282 ACCGCTGGCGCGCATCTACCCCATGCGCGTGTCAAAATTTGTTGGCGGG 1331

411 SerPheCysSerLeuLeuLeuPheIle..... 419
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1332 CGACCTCGCTCGTCATGATTTTATCTCACCTGACCTGCACCTGGTTAA 1381

420 .....CysValGlyArgAsnValLeuThrLeuLeuLeuPheI 432
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1382 CATCATATCATGTGTGTGGCCGAATGGAATCACCATTGCAATA.... 1427

432 leAlaArgAlaPheIleSerGlyGlyPheGlnAlaAlaTyrValTyrThr 448
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1428 .....CAAAATGATCTGCTGGTGAAT 1448

449 ProGluValTyrProThrAlaThrArgAlaLeuGlyLeuGlyThrCysSe 465
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1449 GCTGAGCTGTACCCACATTCGTCAGGAACCTCGGAGTGTGCTGTTC 1498

465 rGlyMetAlaArgValGlyAlaLeuIleThrProPheIleAlaGlnValM 482
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1499 CTCCTGTGTGACATAGGTGGGATAATCACCCCTTCATAGTCTTCAGGC 1548

482 eLeuGluSerSerValTyrLeuThrLeuAlaValTyrSerGlyCysCys 498
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1549 TGAGGGAGGTCTGGCAACCTTGCCCTTCATTTGTTGCGGTGTGGGC 1598

499 LeuLeuAlaAlaLeuAlaSerCysPheLeuProIleGluThrLysGlyGl 515
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1599 CTGCTTGGCGGGAGTGACGCTACTTCTTCCA...GAGACCAAGGGGT 1645

515 yGlyLeuGlnGluSer 520
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1646 CGCTTTGGCAGAGACC 1661

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH57556
seq_documentation_block:
ID AAH57556 standard; cDNA; 4366 BP.
XX
AC
XX
XX
DT 10-SEP-2001 (first entry)

```

```
XX Human brain cell specific cDNA sequence SEQ ID NO:396.
DE
XX Human: tissue specific; diagnosis; brain; heart; skeletal muscle;
KW lung; liver; uterus; ovary; stomach; intestine; kidney; pancreas; ss;
KW metabolic disease; developmental disease; cytostatic; immunomodulatory;
KW neuroprotective; gene therapy; cancer; immunopathology; neuropathology.
XX Homo sapiens.
OS
XX WO200132927-A2.
PN
XX 10-MAY-2001.
XX
XX 02-NOV-2000; 2000WO-US30396.
XX
XX 04-NOV-1999; 99US-0163508.
XX
XX (INCY-) INCYTE GENOMICS INC.
PA
XX Sornasse T, Sellhamer JJ, Watson GA;
PI
XX WPI; 2001-291057/30.
DR
XX
XX New cell and tissue specific polynucleotides useful for diagnosis,
PT prognosis or monitoring of treatments for disorders where the gene is
PT associated with a cancer, immunopathology or neuropathology.
XX
XX Claim 1: Page 309-310; 327pp; English.
XX
XX AAH57161 to AAH57576 represent cell and tissue specific polynucleotide
CC sequences (I). (I) can have cytostatic, immunomodulatory and
CC neuroprotective activities, and can be used in gene therapy. (I) and
CC assays to select DNA molecules, RNA molecules, peptide nucleic acids,
CC mimetics, peptides, proteins, agonists, antagonists, antibodies or
CC their fragments, immunoglobulins, inhibitors, drug compounds and
CC pharmaceutical agents. Expression of (I) in a sample indicates the
CC differentiation of embryonic stem cells into a tissue selected from
CC brain, heart, kidney, liver, lung, skeletal muscle or pancreatic
CC tissues. (I) and (II) are used to produce an expression profile that
CC defines a metabolic or developmental process, treatment, condition,
CC disease or disorder. The gene profile can be used for diagnosis,
CC prognosis or monitoring of treatments and for investigating a
CC predisposition to a disorder where the gene is associated with a
CC cancer, immunopathology or neuropathology.
XX
XX Sequence 4366 BP; 916 A; 1187 C; 1195 G; 1068 T; 0 other;
S0

alignment_scores:
  Quality: 413.00      Length: 710
  Ratio: 1.237        Gaps: 22
  Percent Similarity: 47.042  Percent Identity: 22.254

alignment_block:
US-09-911-667A-2 x AAH57556 ..

Align seg 1/1 to: AAH57556 from: 1 to: 4366

21 GluSerAlaArgSerGluAspThrAlaSerGlyGluHisGluValG1 37
660 GAAGGTGGTCCAGTATGCTACTAGAGGCCATGACGAGGATGATGA 709
37 ntle...GluGlyValHisValClyGluAlaValGlu.....49
710 GATCTATGAAGGGAATATACGGGCATTCGCCGGCAGAGTCTGGGGCA 759
49 .....
760 AAGCGCAGCGGATGCGATGGGGCCGCCCTCGGTGGAGTAAGGGGGGC 809
50 LeuAspGlyAlaAlaValPro.....57
```

```
810 TTGAGTGTGGGGAGGGTCCCTCTGGGGCGGGGGAGGCAACAGCAGC 859
58 .....LysGluPheAlaAsnProThrAspThrPheMetValG 71
860 GAAAGAACGAGAAGAACTGGCCCAA.....CAGTATG 891
71 LuAspAlaValGluAlaTleGlyPheGlyLysPheGlnTrpLysLeuSer 87
892 AAGCCATCCTACGGGAGTGTGGCCACGGCCGCTTCAGTGGACACTGTAT 941
88 ValLeuThrGlyLeuAlaTrpMetAlaAspAlaMetGluMetMetIleLe 104
942 TTGTGCTTGTGCTGGCGTGTGGTGTGAGGTGTGGAGGTCTTTGTGGT 991
104 uSerIleLeuAlaProGlnLeuHisCysGluTrpArgLeuProSerTrpG 121
992 GGGCTTCGTGCTGCCACGGCTGAGAAAGACATGTGCTGTCGACGTCCA 1041
121 InValAlaLeuLeuThrSerValPheValGlyMetMetSerSerSer 137
1042 ACAAGGCGATGTCAGCCCTCATCGTCTACCTGGGCATGATGGTGGAGCC 1091
138 ThrLeuTrpGlyAsnIleSerAspGlnTyrGlyArgLysThrGlyLeuLy 154
1092 TTCTCTGGGGAGGTCTGGCTGACCGGCTGGGTGGGAGGCGAGTCTGTCT 1141
154 sIleSerValLeuTrpThrLeuTyrTyrGlyIleLeuSerAlaPheAlaP 171
1142 CATCTCGCTCTCAGTCACACAGCGCTCTCGCTTCTTCTCATCTTTGTCC 1191
171 rovalTyrSerTrpIleLeuValLeuArgGlyLeuValGlyPheGlyIle 187
1192 AGGTTTACGGCACTTTCCTTCTTCTGCGGCTACTTCTTGGGTGGGATT 1241
188 GlyGly...ValProGlnSerValThrLeuTyrAlaGluPheLeuProme 203
1242 GGAGGTGTCATCCCATTTCTTCTCTCTATTTCTCGAGTTTCTGGCCCA 1291
203 tIysAlaArgAlaLysCysIleLeuLeuIleGluValPheTrpAlaIleG 220
1292 GGAGAAACGAGGGGACATTTGAGTGGCTCTGCATGTTTGGATGATTG 1341
220 LyThrValPheGluValValLeuAlaValPheValMetProSerLeuGly 236
1342 GTGGCGTGTACGAGCTGCTATGGCGCTGGGCCCATCATCCCCACTATGG 1391
237 .....TrpArgTrpLe 240
1392 TGGAGTTTTCAGATGGGTCTTGCCTACCAGTTCCACAGCTGGAGGTCTT 1441
240 uLeuIleLeuSerAlaValProLeuLeuLeuPheAlaValLeuCysPheT 257
1442 CGTCTCTGCTCGGCTTCTTCTCTCTGTTTGGCATTTGGGCTCTGACCA 1491
257 rpLeuProGluSerAlaArgTyrAspValLeuSerGlyAsnGlnGluLys 273
1492 CGCAGCTGAGAGCCCGCTTCTTCTCTAGAGAATGAAAGCATGATGAG 1541
274 AlaIleAlaThrLeuLysArgIleAlaThrGluAsn.....G1 286
1542 GCCTGGATGGTCTGAAGCAGGTTCATATACCAACATCGGAGCCAAAGG 1591
286 yAlaPro.....MetProLeuGlyLysLeuIleLeuSerArgGlnG 300
1592 ACATCTCTGAGCGAGTGTCTTCTCAGTAACCCACATTAAAGACGATTCATCAGG 1641
300 LuAspArg.....302
1642 AGGATGATTTGATGATGATCCAGTCGACACACAGGGACCTGGTACCAGCGC 1691
303 .....GlyLysMetAr 306
```

alignment_scores:		
Quality:	407.00	Length: 556
Ratio:	1.443	Gaps: 17
Percent Similarity:	50.719	Percent Identity: 23.741

alignment_block:

US-09-911-667A-2 x AAC85822

Align seg 1/1 to: AAC85822 from: 1 to: 2121

```
62 AsnProThrAspThrPheMetValGluAspAlaValGluAlaIleG1 78
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
44 GACCCGCCACCACTGCGCATGACCTTCTCGGAGATCCTGGACGGTGTGG 93
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
78 y...PheGlyLysPheGlnTrpLysLeuSerValLeuThrGlyLeuAla 94
   |  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
94 AAGCATGGGCCATTTCACAGTCTCTGCATGATAGCCATACCTGGGCCCTCCCGA 143
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
94 rp.....MetAlaAspAlaMetGluMetMetIleLeuSerIleLeuAla 108
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
144 TCCTCAACATGGCCCAACCAACCTGCTGCAGATCTTCACAGCGCGCAC 193
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
109 ProGlnLeuHisCysGlu.....TrpAr 116
   ||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
194 CTTGTCCACACTGTGCGCGCGGCCCAATGCTTCCACAGGCGCTTGGGT 243
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
116 gLeuPro..... 118
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
244 GCTCCCATGGGCCCAAAATGGGAGCCTGAGAGGTGCTCCCTTTGTAC 293
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
118 ..... 118
294 ATCCGCCCAATGCCAGCTTGCCCAATGACACCCAGAGGCCATGGGCCA 343
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
119 .....Se 119
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
344 TGCTGGATGGGTGGTCTACAAACAGCACCAAGGACTCCATTGTGACAGA 393
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
119 rTrpGlnValAla.....LeuLeuThrServ 128
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
394 GTGGACTTGGTGCACCTCAACAACTGAAGGAGATGGCCCACTCA 443
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
128 alValPheValGlyMetMetSerSerThrLeuTrpGlyAsnIleSer 144
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
444 TCTTCATGGCAGGTACTGATTTGGAGGGCTCGTGTGGAGACCTGTCT 493
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
145 AspGlnTrpGlyArgLysThrGlyLeuLysIleSerValLeuTrpThrLe 161
   ||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
494 GACAGTTTGGCGCGAGGCCCTCTGACCTGCAGCTACCTGCTGTGGC 543
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
161 uTyTrpGlyIleLeuSerAlaPheAlaProValTyTrpTrpIleLeuV 178
   ||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
544 AGCCAGCGGCTCGGGTGCAGCCTTCAGCCCACTTCCCATCTACATGG 593
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
178 alLeuArgGlyLeuValGlyPheGlyIleGlyValProGlnSer... 193
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
594 TCTTCCGCTTCTGTGTGGCTTTGGCATCTCAGGCATATACCCTGAGCAC 643
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
194 ValThrLeuTyAlaGluPheLeuPrometLysAlaArgAlaLysCysI1 210
   ||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
644 GTCATCTGAATGTGGAATGGGTGCTACCCGGATGCGGCCCATCATGTC 693
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
210 eLeuLeuIleGluValPheTrpAlaIleGlyThrValPheGluValVall 227
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
694 GACAGCAGCTCGGTACTGCTACACCTTTGGCCAG..... 727
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
227 euAlaValPheValMetProSerLeuGly.....TrpArg 238
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
728 .....TTCATTCTGCCGGCGCTGGCCTACGCCATCCGCCAGTGGCGT 769
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
239 TrpLeuLeuIleLeuSerAlaValProLeuLeuPheAlaValLeuCy 255
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
770 TGGCTGCAGTAACTGTGTCCATTCCTTCTTCCTTCTTCCTATCATC 819
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
255 sPheTrpLeuProGluSerAlaArgTyTrpValLeuSerGlyAsnGlnG 272
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
820 CTGGTGACACCAAGATGCTCAGCTGGTGGTCTTGTCTGTGGAAGTCCT 869
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
```

```
272 LuLysAlaIleAlaThrLeuLysArgIleAlaThr..... 283
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
870 CGAAGCCCTGAAGATACTCCGGCGGTGGCTCTTCAATGCCAAGAAG 919
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
284 GluAsnGlyAlaPrometProLeuGlyLysLeuIleIleSerArgGlnG1 300
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
920 GAAGAGGGAGAAAGGCTCAGCTTGGAGAGCTCAAACTCAACCTGCAGAA 969
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
300 uAsp.....ArgGlyLysMetArgAspLeuPhe...T 310
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
970 GGAGATCTCTTGGCCCAAGCAAGTACACCCCAAGTGACCTGTTCGGGA 1019
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
310 hrProHisPheArgTrpThrLeuLeuLeuTrpPheIleIleTrpPheSer 326
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1020 TACCCATGCTGCGCGCATGACCTTCTGTCTTCCCTGGCCCTGTGTGT 1069
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
327 AsnAlaPheSerTyTrpGlyLeuValLeuLeuThrGlu..... 340
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1070 ACCGGTTTTGCCTACTATAGTTGGCTATGGGTGGGAAGAATTTGGAGCT 1119
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
341 .....LeuPheGlnAlaGlyAspValCysGlyI 350
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1120 CAACCTCTACATCTCCAGATCATCTTTGGTGGGTGCGATGC..... 1162
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
350 leSerSerArgLysLysAlaValGluAlaLysCysSerLeuAlaCysGlu 366
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1162 ..... 1162
367 TyLeuSerGluGluAspTyTrpMetAspLeuLeuTrpThrLeuSerG1 383
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1162 ..... 1162
383 uPheProGlyValLeuValThrLeuTrpIleIleAspArgLeuGlyArgL 400
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1163 ....CCAGCCAAGTTTCATCACCATCTCTCTTAAGCTACCTGGCGCGC 1208
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
400 yslYstThrMetAlaLeuCysPheValIlePheSerPheCysSerLeu... 415
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1209 ATACCACTCAGGCCG:GCGCTGCTCTGGCAGAGGGGCCCATCTTGGCT 1258
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
416 LeuLeuPheIleCysValGly...ArgAsnValLeuThrLeuLeuLeuPh 431
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1259 CTCACCTTTGTGCCCTTGGACTTGCAGCGGTGAGGACAGTATTGGCTGT 1308
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
431 eIleAlaArgAlaPheIleSerGlyPheGlnAlaAlaTyValTyTr 448
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1309 GTTTGGAAAGGATGCTATCCAGCTCTTCAAGCTGCTCTTCTCTCTACA 1358
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
448 hrProGluValTyTrpThrAlaThrArgAlaLeuGlyLeuGlyThrCys 464
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1359 CAAGTCAATTATACCCACAGTCATCAGGCAACAGGTATGGCGGTAAGT 1408
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
465 SerGlyMetAlaArgValGlyAlaLeuIleThrProPheIleAlaGlnVa 481
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1409 AACCTGTGGACCCCGCTGGGAAGCATGGTGTCCCGCTGGTG...AAAA 1455
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
481 lmetLeuGluSerSerValTyTrpLeuThrLeuAlaValTyTrpSerGlyCys 498
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1456 CACGGGTGAGGTACAGCCCTTCATCCCCAATATCATCTACGGGATCACCG 1505
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
498 ysLeuLeuAlaAlaLeuAlaSerCysPheLeuProIleGluThrLysGly 514
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1506 CCCTCTCGGGGCGAGTGTGCGCTCTCTCTCGCT...GAGACCCCTGAAT 1552
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
515 GlyGlyLeuGlnGluSer 520
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1553 CAGCCCTTGCAGAGAGT 1570
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAAL6697

seq_documentation_block:

ID AAAL6697 standard; cDNA; 3580 BP.

XX

157 aLeuTrpThrLeuTyrTyrcGlyileLeuSerAlaPheAlaProValTyr 173
 589 ATCTCTTGGCGCAGCGAGTGGCTCCAGTGTGCTTCAGCCCGCAGCTC 638
 174 SerTrpIleLeuValLeuArgGlyLeuValcGlyPheGlyIleGlyGlyVa 190
 639 ACTGCTATATGATCTCCGATTCTCTGTGGCTGCAGCATCTCGGGCAT 688
 190 lProGluSer...ValThrLeuTyrAlaGluPheLeuProMetLysAlaA 206
 689 TTCTCTGAGCACCATTATCTTGAATGTGGAAATGGGTACCCACCATCCACGC 738
 206 rgAlaLysCysIleLeuLeuIleGluValPheTrpAlaIleGlyThrVal 222
 739 GGGCCATCTCATCAACAACATATTTCGGTACTCTACACCATTTGGTCAA... 785
 223 PheGluValValLeuAlaValPheValMetProSerLeuGly... 236
 786TTCAATCTCGCTGGCTGGCTACGCTCTCGCTTCCTTCATCT 814
 237TrpArgTrpLeuLeuIleLeuSerAlaValProLeuLeuLeuP 251
 815 TCCTCAGTGGCGCTGGCTACAGTTGTCGCTGTCTGCTGCTTCCTTCATCT 864
 251 heAlaValLeuCysPheTrpLeuProGluSerAlaIaArgTyrAspValLeu 267
 865 TCTCCTTGTCTCTGCTGGGTACCCAGTCCATACGCTGGCTGGTCTG 914
 268 SerGlyAsnGlnGluLysAlaIleAlaThrLeuLysArgIleAlaThr... 283
 915 TCTGGAAATTCACAGAGCTGAAGACACTCAAGCTGTGGGTACGCTT 964
 284GluAsnGlyAlaProMetProLeuGlyLysLeuIleI 296
 965 CAACGGCAAGAAGGAGGAAAGGCTCACTGTGGAGGAGCTGAAGT 1014
 296 leSerArgGlnGluAsp.....ArgGlyLysMetArg 306
 1015 TCAACTTTCAGAGGAGCATCACCTCAGCCAAAGGTCAATATGGCTTATCT 1064
 307 AspLeuPhe...ThrProHisPheArgTrpThrLeuLeuLeuTrpPh 322
 1065 GACTGTTCGAGTGCTCCATCTCGCGCGGTGACCTTCTCTCTCTCT 1114
 322 eileTrpPheSerAsnAlaPheSerTyrTyrcGlyLeuValLeuLeuThrT 339
 1115 GGCCTGGTGTGCTACTGGCTTGTGCTACTACAGTTTGGCTTACGGAGTAG 1164
 339 hrGluLeuPheGlnAlaGlyAspValcysGlyIleSerSerArgLysLys 355
 1165 AAGAAATTT.....GGAGTCAACATC..... 1184
 356 AlaValGluAlaLysCysSerLeuAlaCysGluTyrLeuSerGluGluAs 372
 1185TACATA..... 1190
 372 pTyrMetAspLeuLeuTrpThrLeuSerGluPheProGlyValLeuV 389
 1191CTCCAGATCATCTCGTGGGGTT...GACATTCGCCGCCAAGTCA 1233
 389 alThrLeuTrpIleIleAspArgLeuGlyArgLysLysThrMetAlaLeu 405
 1234 TCACAATCTCTCCATAAGTTATCTGGCGCGGCATCATCTCAGGCGTTC 1283
 406 CysPheValIlePheSerPheCysSerLeu...LeuLeuPheIleCysVa 421
 1284 CTCCTGATCTCGCAGGAGTGGGCATCTCTGGCCCTCATCTTTGTGTCTTC 1333
 421 lGlyArgAsnValLeu...ThrLeuLeuLeuPheIleAlaArgAlaPheI 437
 1334 AGAAATC...AGCTCTTGTGAGAAGACGACGCTGTATTTTGGGAAGGATGCC 1383
 437 leSerGlyGlyPheGlnAlaAlaTyrValTyrThrProGluValTyrPro 453

237TrpArgTrrpleuLeuIleLeuSerAlaValProLeuLeuLeuP 251
||||| ||||| : : : : : : : : : : : : : : : : :
815 TCCTCAGTGGCGGTACAGATTGTCGGTGCTGCCTGCCCTCTCTCATCT 864

251 heAlaValLeuCysPheTrpleuProgluSerAlaAArgTyAspValLeu 267
:
865 TCTCCTTTGTCTCGTGGGTACCAGAGTCCATACCTGGCTGGTGTCTG 914

268 SerGlyAsnGlnGlutylsAlaIleAlaThrLeuLysArgIleAlaThr... 283
||||| :
915 TCTGAAAYTCTCACAGCTCTGAACACACTCCAACGTGTGGCTACTT 964

284gluAsnGlyAlaProMetProLeuGlyLysLeuIleI 296
:
965 CAACGCGAAGAAGGAAGGGAAAAAGCTCACTGTGGAGGAGCTGAAGT 1014

296 leSerArgGlnGluASP.....ArgGlyLysMetArg 306
:
1015 TCAAATTTCAGAGGACACTCACCTCAGCCAAAGTCAAAATATGGCTTATCT 1064

307 AspLeuPhe...ThrProHisPheArgTrpThrThrLeuLeuLeuTrpPh 322
||||| :
1065 GACTGTTCGAGTGTCCATCTCGCGCGGTGACCTTCTGCTCTCTCTCT 1114

322 eileTrpPheSerAsnAlaPheSertyrTyrglyLeuValLeuLeuThrt 339
:
1115 GGCTGGTTTCTACTTGGCTTACTACAGTTTGGCTATGGAGTAG 1164

339 hrGluLeuPheGlnAlaGlyAspValCysGlyIleSerSerArgLysLys 355
:
1165 AAGAATTT.....GGAGTCAACATC..... 1184

356 AlaValGluAlaLysCysSerLeuAlaCysltyrLeuSerGluGluAs 372
:
1185TACATA..... 1190

372 pTyrMetAspLeuLeuTrpThrThrLeuSerGluPheProglyValLeuV 389
:
1191CTCCAGATCATCTCGTGGGTT...GACATCCCGCCCAAGTCA 1233

389 alThrLeuTrpIlelleAspArgLeuGlyArglysLysThrMetalLeu 405
:
1234 TCACAATCTCTCCATAAGTTATCTGGCGCGGCATCACTCAGGGGCTC 1283

406 CysPheValIlePheSerPheCysSerLeu...LeuLeuPheIleCysVa 421
:
1284 CTCCTGATCTCGCAGAGTGGCATCTCTGGCCCTCATCTTTGTGTCTC 1333

421 lGlyArgAsnValLeu...ThrLeuLeuPheIleAlaArgAlaPheI 437
:
1334 AGAATGTAAGTCTTGGAGAACACAGCTGGCTGTATTGGGAAGGATGCC 1383

437 leSerGlyGlyPheGlnAlaAlaTyValTyThrProGluValTyPro 453

```

1384  TGTCTGGGCTCTCTCAGCTGGCTTCTTCTCTACACAAGTGAAGCTCTTACCCT 1433
      :|||::||::||::||::||::|||  |||::|||::|||::|||::|||::|||::|||
454   ThrAlaThrArgAlaLeuGlyLeuGlyThrCysSerGlyMetAlaArgVa 470
      :|||::|||::|||::|||::|||  :|||::|||::|||::|||::|||::|||
1434  ACAGTCTTCAGGCAACAGGTATGGGTATCAGTAAACATATGGGCTCGAGT 1483
      :|||::|||::|||::|||::|||  :|||::|||::|||::|||::|||::|||
470   lGlyAlaLeuIleThrProPheIle.....AlaGlnValM 482
      :|||::|||::|||::|||::|||  :|||::|||::|||::|||::|||::|||
1484  GGAAGTATGATAGCCCCACTGGTGAATAACGGGAGAACTCGACGCCCT 1533
      :|||::|||::|||::|||::|||  :|||::|||::|||::|||::|||::|||
482   eLeuGlnSerSerValrLeuThrLeuAlaValTyrSerGlyCysCys 498
      :|||::|||::|||::|||::|||  :|||::|||::|||::|||::|||::|||
1534  TCATCCCTTAATGTCTATCTTTGGACCATGACTCTACCTGGGAGGC..... 1577
      :|||::|||::|||::|||::|||  :|||::|||::|||::|||::|||::|||
499   LeuLeuAlaAlaLeuAlaSerCysPheLeuProIleGluThrLysGlyI 515
      :|||::|||::|||::|||::|||  :|||::|||::|||::|||::|||::|||
1578  .....AGTGCTGCCTCTCTTCTGTCTTCAGACCCCTCAATCG 1612
      :|||::|||::|||::|||::|||  :|||::|||::|||::|||::|||::|||
515   yGlyLeuGlnGluSerSerHis.....ArgGluTrpGlyGlnGlu 528
      :|||::|||::|||::|||::|||  :|||::|||::|||::|||::|||::|||
1613  GCCCTTACCAGAAACTATCTCGAGGACATACAAGACTGTGTACACGAA 1658
      :|||::|||::|||::|||::|||  :|||::|||::|||::|||::|||::|||

```

OM of: US-09-911-667A-2 to: Issued_Patents_NA:* out_format : pfs
Date: Mar 8, 2002 3:39 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame1.p2n.model -DEV=xlip
-Q=/cgn2_1/USPTO_spool/US09911667/runat_07032002_130318_29553/app_query.fasta_1.615
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DCAALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09911667@cgn1_1.86 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-911-667A-2
Query length: 548
Database: Issued_Patents_NA:*
Database sequences: 351203
Database length: 1,4238999
Search time (sec): 89.480000

score_list:	Sequence	Strd Orig	ZScore	EScore Len	! Documentation
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-501-572-4	1882	1	430.00	884.62	1.3e-41
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-040-444-4	1882	1	430.00	884.62	1.3e-41
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-501-572-6	1886	1	426.00	875.86	3.9e-41
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-040-444-6	1896	1	426.00	875.86	3.9e-41
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-501-572-5	1885	1	425.50	874.85	4.4e-41
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-040-444-5	1885	1	425.50	874.85	4.4e-41
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-647-397-1	2102	1	401.00	820.34	4.8e-38
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-1	4411529	1	230.00	347.79	1.0e-11
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-964-127-1	2460	1	203.50	390.56	4.2e-14
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2	4403765	1	166.00	209.22	0.0005
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-1	4411529	1	158.00	191.87	0.0049
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-591-025-8	2592	1	154.50	283.75	3.7e-08
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-125-468-1	30001	1	151.50	244.52	5.7e-06
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-474-933-1	30001	1	151.50	244.52	5.7e-06
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-928-692-9	3000	1	147.00	265.55	3.8e-07
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-031-392-1	2343	1	145.00	264.53	4.4e-07
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-299-549-1	2343	1	145.00	264.53	4.4e-07
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-994-035C-4	7198	1	142.00	243.03	6.9e-06
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-494-907-13	1269	1	138.00	257.57	1.1e-06
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-494-907-2	1269	1	138.00	257.57	1.1e-06
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-494-907-2	5076	1	138.00	239.03	1.2e-05
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-494-907-2	5076	1	138.00	239.03	1.2e-05
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-494-907-4	6170	1	138.00	236.42	1.6e-05
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-494-907-4	6170	1	138.00	236.42	1.6e-05
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-494-907-3	6387	1	138.00	235.96	1.7e-05
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-494-907-3	6387	1	138.00	235.96	1.7e-05
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-964-127-5	5390	1	133.00	229.57	3.9e-05
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-964-127-5	1411	1	133.00	245.32	5.1e-06
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-882-704A-4	7742	1	131.50	219.31	0.0001
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-07-925-695-3	9589	1	129.00	222.83	9.2e-05
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-07-925-695-1	9589	1	129.00	211.04	0.0004
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-07-925-695-2	9589	1	129.00	211.04	0.0004
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-964-127-3	1490	1	127.00	231.60	3.0e-05
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-931-999-4	6755	1	125.50	208.14	0.0006
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-850-880-11	19182	1	119.50	180.11	0.0192
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-944-916-11	19182	1	119.50	181.20	0.0192
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-024-020B-1	5977	1	116.50	190.29	0.0060
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-024-020B-2	6007	1	116.50	190.22	0.0060
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-338-702-7	6513	1	116.50	189.14	0.0069

/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-337-339-7 + 116.50 189.14 0.0069 6513
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-724-095-7 + 116.50 189.14 0.0069 6513
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-14262-7 + 116.50 189.14 0.0069 65
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-14378-7 + 116.50 189.14 0.0069 65
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-024-020B-7 + 116.50 189.06 0.0070 6556

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-501-572-4

seq_documentation_block:

: Sequence 4, Application US/08501572

: Patent No. 6063623

: GENERAL INFORMATION:

: APPLICANT: Koepsell, Hermann

: APPLICANT: Grundeman, Dirk

: TITLE OF INVENTION: Transport of Cationic Xenobiotics and/or Pharmaceuticals.

: TITLE OF INVENTION: Transport of Cationic Xenobiotics and/or Pharmaceuticals.

: TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.

: NUMBER OF SEQUENCES: 6

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Finnegan,Henderson,Farabow,Garrett & Dunner

: STREET: 1300 I Street, N.W., Suite 700

: CITY: Washington

: STATE: D.C.

: COUNTRY: USA

: ZIP: 20005-3315

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: Patent In Release #1.0, Version #1.30

: CURRENT APPLICATION DATA:

: FILING DATE:

: CLASSIFICATION: 424

: ATTORNEY/AGENT INFORMATION:

: NAME: Toohey, Kimberlin M

: REGISTRATION NUMBER: 35,391

: REFERENCE/DOCKET NUMBER: 02481.1453-00000

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: (202)408-4000

: TELEFAX: (202)408-4400

: INFORMATION FOR SEQ ID NO: 4:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 1882 base pairs

: TYPE: nucleic acid

: STRANDEDNESS: single

: TOPOLOGY: linear

: MOLECULE TYPE: DNA (genomic)

: US-08-501-572-4

alignment_scores:

Quality: 430.00 Length: 441

Ratio: 1.713 Gaps: 14

Percent Similarity: 56.916 Percent Identity: 29.478

alignment_block:

US-09-911-667A-2 x US-08-501-572-4

Align seg 1/1 to: US-08-501-572-4 from: 1 to: 1882

119 serTrpGlnValAlaLeuLeuThrSerValPheValGlyMetMetSe 135

119 serTrpGlnValAlaLeuLeuThrSerValPheValGlyMetMetSe 135

473 GCCTGGAAGTGGACCTTTTCAGTCTCTGCTGAACTTGGCTTCTCCT 522

135 rSerSerThrLeuTrpGlyAsnIleSerAspGlnTyrClyArgLysThrG 152

523 GGGCTCCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 572

152 LyLeuLysIleSerValLeuTrpThrLeuTyrTyrGlyIleLeuSerAla 168

573 GTCTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 622

449 ProGluValTyrProThrAlaThrArgAlaLeuGlyLeuGlyThrCys 465
1400 GCTGAGCTGTACCTACATTCATCAGGAATCTTGGATGATGATGCTC 1449
465 rGlyMetAlaGValGlyAlaLeuLeuThrProPheIleAlaGlnValM 482
1450 TGCCTGTGTGACCTGGGTGGGATCTTACCCCTTCATGCTGTTGAGGC 1499
482 etLeuGluSerSerValTyrLeuThrLeuAlaValTyrSerGlyCys 498
1500 TGATGAAGTTTGGCAAGCCTGCCTCTCATTTTGTGGGTTTGGGC 1549
499 LeuLeuAlaLeuAlaSerCysPheLeuProIleGluThrLysGlyGI 515
1550 CTGACTGCTGGGCCATGACTCTCTCTCCCA...GAGACCAAGGTGT 1596
515 yGlyLeuGlnGluSer.....Ser.HisArgGluTrpGly.GlnGluMe 529
1597 GGCTTTGCCTGAGACTATTCAACAGCAGAACCTGGGAGGAGGAAT 1646
529 tValGlyArgGlyMethis 535
1647 CAAAGGCCAAAGAAACAC 1665

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:us-09-040-444-4

seq_documentation_block:
; Sequence 4, Application US/09040444
; Patent No. 6063766
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein Which Effects The
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics And/or Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan,Henderson,Parabow,Garrett & Dunner, L.L.P.
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,444
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1882 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-040-444-4

alignment_scores:
Quality: 430.00 Length: 441
Ratio: 1.713 Gaps: 14

169 PheAlaProValTyrSerTrpIleLeuValLeuArgGlyLeuValGlyPh 185
623 GTGCCCCAGACTATACATCATCTTGTCTCTTTCGGCTGCTGAGGCAT 672
185 e...GlyIleGlyValProGlnSerValThrLeuTyrAlaGluPheL 201
673 GGTACGACGAGGAGCTGGGTGCTCGGCTATACCTTGCATCAGAGTTG 722
201 euProMetLysAlaArgAlaLysCysIleLeuLeuIleGluValPheTrp 217
723 TCGGCTCTGGCTACAGGAGAACACGCGCCATTTGTACCATGAGCTTC 772
218 AlaIleGlyThrValPheGluValValLeuAlaValPheValMetProSe 234
773 ACATGGGGCTAGTGGGGCTTGGGGGTGGCC...TATGCCATTCAGA 819
234 rLeuGlyTrpArgTrpLeuLeuIleLeuSerAlaValProLeuLeuLeuP 251
820 C.....TGGCGCTGGCTCCAGCTAGCTGTGCTCCCTGCTACCTCTCT 863
251 heAlaValLeuLysPheTrpLeuProGluSerAlaArgTyrAspValLeu 267
864 TCCTGCTGTATTACTGTTTGTCCAGAAATCCCGCCGCTGCTGCTGCTCC 913
268 SerGlyAsnGlnGluLysAlaIleAlaThrLeuLysArgIleAlaThrGI 284
914 CAGAGACACACCGGAGCTGTCCAGGATAATGAGCAAAATTCACAGAA 963
284 uAsnGlyAlaProMetProLeuGlyLysLeuIleIleSerArgGlnGlu 301
964 GAACGGGAAGTGGCTGCTGCTGACCTGAAGATGCTCTGCTGAGGAGG 1013
301 sPArgGlyLysMetArg.....AspLeuPhe...ThrPro 311
1014 ATGCCCTCAGAAAGGAGTCTTCGTTGGCGGACCTGTTCCGCACCTCC 1063
312 HisPheArgTrpThrLeuLeuLeuTrpPheIleTrpPheSerAsnAl 328
1064 AACCTGAGGAGACACACCGCTCATCTGATGATATGTTCTCTGTGTC 1113
328 aPheSerTyrTyrGlyLeuValLeuThrThrGluLeuPheGlnAlaG 345
1114 TGTGCTGTACAGGCTCTCATCATG..... 1138
345 lyAspValCysGlyIleSerSerArgLysLysAlaValGluAlaLysCys 361
1139CACGTGGGAGGCACACAGGGGCCAACCTC 1165
362 SerLeuAlaCysGluTyrLeuSerGluGluAspTyrMetAspLeuLeuTr 378
1166TACCTGGACTCTTTTIA 1182
378 pThrThrLeuSerGluPheProGlyValLeuValThrLeuTrpIleLea 395
1183 TTCTCTCTGGGGAATCCCGCGGCTTCATCATCTGCTGCTACCATG 1232
395 sPArgLeuGlyArgLysLysThrMetAlaLeuCys....PheValIlePhe 410
1233 ACCGATTTGGCGGATCTACCAATAGCGGCTCGAATCTGTGTACGGGG 1282
411 SerPheCysSerLeuLeuLeuPheIle..... 419
1283 GCAGCTGCTCTCTCATGATCTTTATCCCGCATGAGCTGCATGCTTCAA 1332
420CysValGlyArgAsnValLeuThrLeuLeuLeuPheI 432
1333 CGTTACCTCGCTGCTGTGGCGGTATGGGGCCACCATTTGCTG.... 1378
432 leAlaArgAlaPheIleSerGlyGlyPheGlnAlaAlaTyrValTyrThr 448
1379CAGATGGTCTGCTGCTGTAAC 1399

Tue Mar 12 15:54:27 2002

Percent Similarity: 56.916 Percent Identity: 29.478

alignment_block:

US-09-911-667A-2 x US-09-040-444-4 ..

Align seg 1/1 to: US-09-040-444-4 from: 1 to: 1882

```

119 SerTrpGlnValAlaLeuLeuThrSerValValPheValGlyMetMet 135
    ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
473 GCTTGAAGAGTGGACCTTTTCAGTCTCTGTGAACTTGGCTTCTCT 522
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
135 rSerSerThrLeuTrpGlyAsnIleSerAspGlnIleGlyArgIleThr 152
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
523 GGGTCCCTGGTGTGGTGTACATTCAGACAGGTTGGCCGTAAGCTCT 572
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
152 lLeuIleSerValLeuTrpThrLeuTrpThrGlyIleLeuSerAla 168
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
573 GTCTTGGTGACCGCTGGTGCATCTGTGCTCCGGTGTCTAACAGCG 622
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
169 PheAlaProValTyrSerTrpIleLeuValLeuArgGlyLeuValGly 185
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
623 GTGCCCCAGACATACATCATCTTGTCTTTCGGCTGTGTCAGGCAT 672
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
185 e...GlyIleGlyValProGlnSerValThrLeuTyrAlaGluPhe 201
    ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
673 GGTGACGAGGCGAGCTGGTGTGCGGCTATACCTTGATCAGAGTTG 722
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201 euProMetLysAlaAlaGlyCysIleLeuLeuIleGluPheTrp 217
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
723 TCGCTCTGCTACAGGAGACGAGCGCCATTTGTACCATGAGCTTC 772
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
218 AlaIleGlyThrValPheGluValAlaLeuAlaValPheValMetPro 234
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
773 ACAGTGGGGCTAGTGGGGCTTGGGGGGTGGCC...TATGCCATTCC 819
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 rLeuGlyTrpArgTrpLeuIleLeuSerAlaValProLeuLeuLeu 251
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
820 C.....TGGCGCTGGCTCCAGCTAGTGTCTGCTCCCTACCTTCT 863
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
251 heAlaValLeuCysPheTrpLeuProGlnSerAlaArgTyrAspVal 267
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
864 TCCTGCTGTTACTGTTGTCCCAAGATCCCGCGGTGGCTGTGTCC 913
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
268 SerGlyAsnGlnGluLysAlaIleAlaThrLeuLysArgIleAlaThr 284
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
914 CAGAAGAACACCGAGCTGTCCAGATATGAGCAATTCACACAGAA 963
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
284 uAsnGlyAlaProMetProLeuGlyLysLeuIleLeuSerArgGln 301
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
964 GAACGGGAAGGTGCTCTGCTGACCTGAAGATGCTCTGCTTGGAGG 1013
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
301 spArgGlyLysMetArg.....AspLeuPhe...ThrPro 311
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1014 ATGCTCTAGAAAGCGAGTCTTGGCTTGGCGACCTGTTCGGACATCC 1063
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
312 HisPheArgTrpThrLeuLeuLeuTrpPheIleTrpPheSerAsn 328
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1064 AACCTGAGAAACACACCGTCTATCTATGTTCTCTCTGTGC 1113
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
328 aPheSerTyrTrpGlyLeuValLeuLeuThrThrGluLeuPheGln 345
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1114 TGTGCTGACAGGGTCTCATCATG..... 1138
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
345 lAspValCysGlyIleSerArgLysLysAlaValGluAlaLysCys 361
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1139 .....CACGTGGAGGCCACAGGGGCCAACCTC 1165
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
362 SerLeuAlaCysGluTyrLeuSerGluAspTyrMetAspLeuLeu 378
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1166 .....TACCTGGACTTCTTTA 1182
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
378 pThrThrLeuSerGluPheProGlyValLeuValThrTrpIleAla 395
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

1183 TTCCTCTCTGGTGAATTTCCCGCGCCTTCATCATCTCTGGTCCACAT 1232
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
395 spArgLeuGlyArgLysLysThrMetAlaLeuCys...PheValIlePhe 410
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1233 ACCGCAATGGCGCATCTACCCAATAGCGCCTCGAATCTGGTGACGGG 1282
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
411 SerPheCysSerLeuLeuLeuPheIle..... 419
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1283 GCAGCTGCTCTCTCATGATCTTTATCCCGCATGAGCTGCATGTTGAA 1332
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
420 .....CysValGlyArgAsnValIleuThrLeuLeuLeuPheI 432
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1333 CGTTACCTCGCTGCTTGGCGGTATGGGGCCACCATTTGTGCTG... 1378
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
432 leAlaArgAlaPheIleSerGlyGlyPheGluAlaAlaTyrValThr 448
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1379 .....CAGATGCTCTGCTGGTGAAC 1399
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
449 ProGluValTyrProThrAlaThrArgAlaLeuGlyLeuGlyThrCys 465
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1400 GTGAGCTGTACCTACATTCATCAGGAATCTTGGGATGATGATGCTC 1449
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
465 rGlyMetAlaArgValGlyAlaLeuIleThrPropheIleAlaGlnVal 482
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1450 TGCCCTGTGTGACCTGGTGGGTCTTCACTCCCTTCATGTTTCAGGC 1499
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
482 etLeuGluSerSerValTyrLeuThrLeuAlaValTyrSerGlyCys 498
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1500 TGATGGAAGCTTTGGCAAGCCTGCCCCCTCATTTTGTGGGGTTGGGC 1549
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
499 LeuLeuAlaAlaLeuAlaSerCysPheLeuProIleGluThrLysGly 515
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1550 CTGACTGTGGGGCCATGACTCTTCTTCTCCA...GAGACCAAGGTGT 1596
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
515 yGlyLeuGlnGluSer.....SerHisArgGluTrpGlyGlnGluMe 529
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1597 GGCTTTGCTGTGACATTTGAAGAGCAGAGAACCTGGGGAGGAGAAAT 1646
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
529 tValGlyArgGlyMetHis 535
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1647 CAAAGGCCAAAGAAACAC 1665
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-501-572-6
seq_documentation_block:
; Sequence 6, Application US/08501572
; Patent No. 6063623
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein which effects the
; TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
; TITLE OF INVENTION: DNA sequences encoding it and their use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/501,572
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Toohey, Kimberlin M

```



```
;
; REGISTRATION NUMBER: 35,391
; REFERENCE/DOCKET NUMBER: 02481.1453-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1896 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
;
US-08-501-572-6

alignment_scores:
  Quality: 426.00      Length: 429
  Ratio: 1.760        Gaps: 14
  Percent Similarity: 56.410  Percent Identity: 29.371

alignment_block:
US-09-911-667A-2 x US-08-501-572-6
..
Align seg 1/1 to: US-08-501-572-6 from: 1 to: 1896

119 SerTrpGlnValAlaLeuLeuThrSerValValPheValGlyMetMetSe 135
|||||:|||||:|||||: ||| ||| |||||:|||||:
580 TCCTGGATGTTGGACCTATTCCAGCTACAGTCAGTGAATGATGAGTCTTTAT 629
|||||:|||||:|||||: ||| ||| |||||:|||||:
135 rSerSerThrLeuTrpGlyAsnIleSerAspClnTrpGlyArgLysThrG 152
|||||:|||||:|||||: ||| ||| |||||:|||||:
630 TGGCTCTATGATGATCGCTACATAGCAGACAGGTTTGGCCGTAAGCTCT 679
|||||:|||||:|||||: ||| ||| |||||:|||||:
152 ILeuLysIleSerValLeuTrpThrLeuTrpGlyIleLeuSerAla 168
|||||:|||||:|||||: ||| ||| |||||:|||||:
680 GCCTCTAACTACAGCTCTCAATAATGCTGCAGCTGGAGTCTCTATGCC 729
|||||:|||||:|||||: ||| ||| |||||:|||||:
169 PheAlaProValTySerTrpIleLeuValLeuArg.....GlyLe 182
|||||:|||||:|||||: ||| ||| |||||:|||||:
730 APTTCCCAACCTATACGTGGATGTTAATTTTCGCTTAATCCAGGACT 779
|||||:|||||:|||||: ||| ||| |||||:|||||:
182 uVal.....GlyPheGlyIleGlyGlyValProGlnSerValThrL 196
|||||:|||||:|||||: ||| ||| |||||:|||||:
780 GGTGCAAAAGCAGGCTGGTTAATAGGCTACATC.....C 814
|||||:|||||:|||||: ||| ||| |||||:|||||:
196 euTyAlaGluPheLeuProMetLysAlaArgAlaLysCysIleLeuLeu 212
|||||:|||||:|||||: ||| ||| |||||:|||||:
815 TGATTACAGAATTTGTTGGGGGAGATATCGGAGAACAGTGGGGATTTT 864
|||||:|||||:|||||: ||| ||| |||||:|||||:
213 IleGluValPheTrpAlaIleGlyThrValPheGluValValLeuAla... 228
|||||:|||||:|||||: ||| ||| |||||:|||||:
865 TACCAAGTTGGCTATACAGTTGGGCTC.....CTGGTGCTAGCTGG 905
|||||:|||||:|||||: ||| ||| |||||:|||||:
229 ....ValPheValMetProSerLeuGlyTrpArgTrpLeuIleLeuS 244
|||||:|||||:|||||: ||| ||| |||||:|||||:
906 GGTGGCTTACGCACCTTCCTCAC.....TGGAGGTGGTTGCAGTTCCAC 949
|||||:|||||:|||||: ||| ||| |||||:|||||:
244 erAlaValProLeuLeuLeuPheAlaValLeuLysCysPheTrpLeuProGlu 260
|||||:|||||:|||||: ||| ||| |||||:|||||:
950 TTGCTCTGCCCAACTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 999
|||||:|||||:|||||: ||| ||| |||||:|||||:
261 SerAlaArgTyAspValLeuSerGlyAsnGlnGluLysAlaIleAlaTh 277
|||||:|||||:|||||: ||| ||| |||||:|||||:
1000 TCTCCAGGTGGCTGATCTCCAGAAATAAGAAATGCTGAAGCCATGAGAAT 1049
|||||:|||||:|||||: ||| ||| |||||:|||||:
277 rLeuLysArgIleAlaThrGluAsnGlyAlaProMetProLeuGlyLysL 294
|||||:|||||:|||||: ||| ||| |||||:|||||:
1050 CATTAAGCACATCGCAAGAAATAATGGAATATCTTACCCGCCCTCCCTTC 1099
|||||:|||||:|||||: ||| ||| |||||:|||||:
294 euIleSerArgGlnGluAspArgGlyLys.....Met 305
|||||:|||||:|||||: ||| ||| |||||:|||||:
1100 AGCGCCTCAGACTTGAAGAGGAAACTGCAAGAAATTAACCCCTTCATTT 1149
|||||:|||||:|||||: ||| ||| |||||:|||||:
```

```
306 ArgAspLeuPhe...ThrProHisPheArgTrpThrThrLeuLeuLeuTr 321
|||||:|||||:|||||: ||| ||| |||||:|||||:
1150 CTTGACTTGGTCAGAACTCCTCAGATAAGGAACATATCATATGAT 1199
|||||:|||||:|||||: ||| ||| |||||:|||||:
321 pPheIleTrpPheSerAsnAlaPheSerTyTyTyGlyLeuValLeuLeu 338
|||||:|||||:|||||: ||| ||| |||||:|||||:
1200 GTACAACGTGGTTACGAGCTCTGCTCTACACAGGCGCTCATCATG.... 1245
|||||:|||||:|||||: ||| ||| |||||:|||||:
338 hrThrGluLeuPheGlnAlaGlyAspValCysGlyIleSerSerArgLys 354
|||||:|||||:|||||: ||| ||| |||||:|||||:
1246 .....CACATGGGCTTGCAGGTGAC..... 1266
|||||:|||||:|||||: ||| ||| |||||:|||||:
355 LysAlaValGluAlaLysCysSerLeuAlaCysGluTyLeuSerGluCl 371
|||||:|||||:|||||: ||| ||| |||||:|||||:
1267 .....AA 1268
|||||:|||||:|||||: ||| ||| |||||:|||||:
371 uAspTyMetAspLeuLeuTrpThrThrLeuSerGluPheProGlyValL 388
|||||:|||||:|||||: ||| ||| |||||:|||||:
1269 TATCTACCTGGATTCTTCTACTCTGCCCTGGTTGAATCCAGCGTGCCT 1318
|||||:|||||:|||||: ||| ||| |||||:|||||:
388 euValThrLeuTrpIleIleAspArgLeuGlyArgLysLysThrMetAla 404
|||||:|||||:|||||: ||| ||| |||||:|||||:
1319 TCATGATCATCTCATTTATCGACGCGATCGGACGCGTTACCGTTGGGCT 1368
|||||:|||||:|||||: ||| ||| |||||:|||||:
405 LeuCys.....PheValIlePheSerPheCysSerLeuLeuLeuPheIle... 419
|||||:|||||:|||||: ||| ||| |||||:|||||:
1369 GCATCAAAATATGGTTGACGGGCGACCGCTGT..GGCCTCAGTTTTTATACC 1418
|||||:|||||:|||||: ||| ||| |||||:|||||:
420 .....CysValGlyArgAsnV 425
|||||:|||||:|||||: ||| ||| |||||:|||||:
1419 TGGTGATCTACAATGGCTAAAAATTTATTTCTCATGCTTGGGACAAATGG 1468
|||||:|||||:|||||: ||| ||| |||||:|||||:
425 alLeuThrLeuLeuLeuPheIleAlaArgAlaPheIleSerGlyGlyPhe 441
|||||:|||||:|||||: ||| ||| |||||:|||||:
1469 GGATCACAATG.....GCCTAT 1485
|||||:|||||:|||||: ||| ||| |||||:|||||:
442 GlnAlaAlaTyValTyThrProGluValTyProThrAlaThrArgAl 458
|||||:|||||:|||||: ||| ||| |||||:|||||:
1486 GAGATAGTCTGCCGTGGTCAATGCTGAGCTGACCCACATTCATTAGGAA 1535
|||||:|||||:|||||: ||| ||| |||||:|||||:
458 aLeuGlyLeuGlyThrCysSerGlyMetAlaArgValGlyAlaLeuIleT 475
|||||:|||||:|||||: ||| ||| |||||:|||||:
1536 TCTTGGCGTCCACATCTGTTCTCAATGTGTGACATTTGGTGCATCATCA 1585
|||||:|||||:|||||: ||| ||| |||||:|||||:
475 hrProPheIleAlaGlnValMetLeuGluSerSerValTyLeuThrLeu 491
|||||:|||||:|||||: ||| ||| |||||:|||||:
1586 GCGCATCTCTGGTCTACCGGCTCACTAACATCTGGCTTGAGCTCCCGCTG 1635
|||||:|||||:|||||: ||| ||| |||||:|||||:
492 AlaValTySerGlyCysCysLeuLeuAlaAlaLeuAlaSerCysPheLe 508
|||||:|||||:|||||: ||| ||| |||||:|||||:
1636 ATGTTTTTCGGCGTACTTGGCTTGGTTGCTGGAGCTCTGGTCTGTGCT 1685
|||||:|||||:|||||: ||| ||| |||||:|||||:
508 uProIleGluThrLysGlyGlyGlyLeuGlnGluSer 520
|||||:|||||:|||||: ||| ||| |||||:|||||:
1686 TCCA...GAACTAAAGGAAAGCTTTGCCCTGAGACC 1719
|||||:|||||:|||||: ||| ||| |||||:|||||:

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-040-444-6

seq_documentation_block:
; Sequence 6, Application US/09040444
; Patent No. 6063/66
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein which effects the
; TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
; TITLE OF INVENTION: DNA sequences encoding it and their use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
```

```

; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,444
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P.
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1896 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-040-444-6

alignment_scores:
  Quality: 426.00      Length: 429
  Ratio: 1.760        Gaps: 14
  Percent Similarity: 56.410      Percent Identity: 29.371

alignment_block:
US-09-911-667a-2 x US-09-040-444-6 ..
Align seg 1/1 to: US-09-040-444-6 from: 1 to: 1896

119 SerTrpGlnValAlaLeuLeuThrSerValValPheValGlyMetMetSe 135
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
580 TCCTGGATGTTGGACCTATCCAGTCATCAGTGAATGTAGGATCTTTAT 629

135 rSerSerThrLeuTrpGlyAsnIleSerAspGlnTyrGlyArgLysThrG 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
630 TGGCTCTATGATGTCGCTACATAGCAGACAGGTTTGGCGCTAAGCTCT 679

152 LyLeuLysIleSerValLeuTrpThrLeuTyrTyrGlyIleLeuSerAla 168
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
680 GCCTCTAACTACAGTCTCTAATAAGTCTGCAGCTGGAGTCTCATGGCC 729

169 PheAlaProValTyrSerTrpIleLeuValLeuArg.....GlyLe 182
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
730 ATTTCCCAACCTATAGCTGATGCTTAATTTTCGCTTAATCCAAAGACT 779

182 uVal.....GlyPheGlyIleGlyValProGlnSerValThrL 196
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
780 GGTGAGCAACAGCGCTGGTTAATAGGCTACATC.....C 814

196 euTyrAlaGluPheLeuProMetLysAlaArgAlaLysCysIleLeuLeu 212
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
815 TGATTACAGAATTTGTTGGGGGAGATATCGAGAACAGTGGGATTTT 864

213 IleGluValPheTrpAlaIleGlyThrValPheGluValValLeuAla.. 228
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
865 TACCAAGTTGCCTATACAGTTGGGCTC.....CTGGTGTAGCTGG 905

229 ...ValPheValMetProSerLeuGlyTyrArgTrpLeuLeuIleLeuS 244
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
906 GGTGGCTTACGCATCTCTAC.....TGGAGGTGGTTGCAGTTCACAG 949

244 erAlaValProLeuPheAlaValLeuCysPheTrpLeuProGlu 260
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

950 TTGCTCTGCCCAACTTCTTCTTCTTCTATTACTTGTGTCATACCTGAG 999

261 SerAlaArgTyrAspValLeuSerGlyAsnGlnGlnLysAlaIleAlaTh 277
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1000 TCTCCAGGTGGTGTCTCCCAAGTAAGAATGCTGAAGCATGAGAAT 1049

277 rLeuLysArgIleAlaThrGluAsnGlyAlaProMetProLeuGlyLysL 294
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1050 CATTAAGCACATCCCAAGAAAAATGGAATCTCTACCCGCCCTCCCTTC 1099

294 euIleIleSerArgGlnGluAspArgGlyLys.....Met 305
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1100 AGCGCTTGAGACTTGAAGAGGAAATGCGAAGAAATGAACCCCTTCATT 1149

306 ArgAspLeuPhe...ThrProHisPheArgTrpThrLeuLeuLeuTr 321
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1150 CTTGACTTGGTCAGAACCTCCACAGTAAAGAAACATCTATGATATTGAT 1199

321 pPheIleTrpPheSerAsnAlaPheSerTyrTyrGlyLeuValLeuLeu 338
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1200 GTACAACCTGGTTACAGAGCTCTGTCTTACCAGGCGCTCATCATG... 1245

338 hrThrGluLeuPheGlnAlaGlyAspValCysGlyIleSerSerArgLys 354
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1246 .....CACATGGGCTTTCAGAGTGAC..... 1266

355 LysAlaValGluAlaLysCysSerLeuAlaCysGluTyrLeuSerGluG 371
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1267 .....AA 1268

371 uAspTyrMetAspLeuLeuTrpThrLeuSerGluPheProGlyValL 388
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1269 TATCTAGTTGGATTCTTCTACTCTGCGCTGGTTGAATTCGAGCTGCCT 1318

388 euValThrLeuTrpIleIleAspArgLeuGlyArgLysLysThrMetAla 404
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1319 TCATGATCATCTCATATACGACCGCATCGGACCGCTTACCCTTGGGCT 1368

405 LeuCys...PheValIlePheSerPheCysSerLeuLeuLeuPheIle... 419
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1369 GCATCAAAATATGGTTGCGAGGGCAGCCTGTCTGGCCTCAGTTTATACC 1418

420 .....CysValGlyArgAsnV 425
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1419 TGGTCATCTACAATGGCTTAAATAATATTATCTCATGCTTGGGAAGAATGG 1468

425 alLeuThrLeuLeuPheIleAlaArgAlaPheIleSerGlyGlyPhe 441
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1469 GGATCACAAATG.....GCCTAT 1485

442 GlnAlaAlaTyrValTyrThrProGluValTyrProThrAlaThrArgAl 458
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1486 GAGATAGTCTGCCTGGTCAATGCTGAGTGTACCCACACATTCATTAGGAA 1535

458 alLeuGlyLeuGlyThrCysSerGlyMetAlaArgValGlyAlaLeuIle 475
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1536 TCTTGGGCTCCACATCTGTTCTCAATGTGTGACATTTGGTCATCATCA 1585

475 hrPropheIleAlaGlnValMetLeuGluSerSerValTyrLeuThrLeu 491
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1586 CGCCATCTCTGGTCTACCGGCTCACTAACATCTGGCTTGAGCTCCCGCTG 1635

492 AlaValTyrSerGlyCysLeuLeuAlaLeuAlaLeuAlaSerCysPheLe 508
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1636 ATGTTTTTCGGCGTACTTGGCTTGGTGTGGAGTCTGGTGTGCTTGTGCT 1685

508 uProIleGluThrLysGlyGlyLeuGlnGluSer 520
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1686 TCCA...GAAACTAAGGGAAGCTTTGCCTGAGACC 1719

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-501-572-5
seq_documentation_block:

```

```

: Sequence 5, Application US/08501572
: Patent No. 6063623
: GENERAL INFORMATION:
: APPLICANT: Koepsell, Hermann
: APPLICANT: Grundeman, Dirk
: APPLICANT: Gorboulev, Valentin
: TITLE OF INVENTION: Transport protein which effects the
: TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
: TITLE OF INVENTION: DNA sequences encoding it and their use.
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
: STREET: 1300 I Street, N.W., Suite 700
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-3315
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/501,572
: FILING DATE:
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Toohey, Kimberlin M
: REGISTRATION NUMBER: 35,391
: REFERENCE/DOCKET NUMBER: 02481.1453-00000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202)408-4000
: TELEFAX: (202)408-4400
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1885 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-501-572-5

```

```

alignment_scores:
  Quality: 425.50      Length: 421
  Ratio: 1.766         Gaps: 11
  Percent Similarity: 57.245  Percent Identity: 30.166

alignment_block:
  US-09-911-667A-2 x US-08-501-572-5
  Align seg 1/1 to: US-08-501-572-5 from: 1 to: 1885
  119 SerTrpGlnValAlaLeuLeuThrSerValValPheValGlyMetMetSe 135
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  505 TCCTGGAAGCTGGACCTCTTTCAGTCTGTTGAATGCGGCTCTCTCTT 554
  135 rSerSerThrLeuTrpGlyAsnIleSerAspGlnTyrGlyArgLysThrG 152
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  555 TGCTCTCTCGGTGGCTGCTTTCAGACAGGTTCGCGGCTGCTGCTGCTG 604
  152 lLeuLysIleSerValLeuTrpThrLeuTyrTyrGlyIleLeuSerAla 168
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  605 GTCTCTGGGAAGCTGCTGCTCAACGCGGCTGCGGCGTCTGCTGCTGCTG 654
  169 PheAlaProValTyrSerTrpIleLeuValLeuArgGlyLeuValGlyPh 185
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  655 TTCTCGGCCCAACATACATGCTCATGCTCTCTCCGCTGCTGCTGCTGCTG 704
  185 e...GlyIleGlyValProGlnSerValThrLeuTyrAlaGluPheL 201
  :...:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  705 GGTGACGACGGCACTGGATGCTGCTACACCTTAATCACAGAAATTG 754

```

```

201 euProMetLysAlaArgAlaLysCysIleLeuLeuIleGluValPheTrp 217
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
755 TTGGCTCGGGCTCCAGAGAAGCGGTGGCGATCATGTACAGCATGGCTTC 804
218 AlalleGlyThrValPheGluValValLeuAlaValPheValMetProse 234
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
805 ACGGTGGGGCTGGTGGCGCTTACCGGGCTGGCC...TAGGCCCTGCTCA 851
234 rLeuGlyTyrArgTrpLeuLeuIleLeuSerAlaValProLeuLeuLeuP 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
852 C.....TGGCGCTGGCTGCAGTGCAGTCTCCCTGCCACCTTCCTCT 895
251 heAlaValLeuCysPheTrpLeuProGluSerAlaArgTyrAspValLeu 267
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
896 TCCTGCTCTACTACTGTGTGCGCGAGTCCCTCGGTGGCTGTATCA 945
268 SerGlyAsnGlnGluLysAlaIleAlaThrLeuLysArgIleAlaThrGl 284
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
946 CAAAAAGAAACACTGACGCAATAAAGATAATGACACCATCGCTCAAAA 995
284 uAsnGlyAlaProMetProLeuGlyLysLeuIleIleSerArgGlnGluA 301
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
996 GAATGGGAAGTTGCTCTGCTGATTTAAAGATGCTTCCCTCGAAGAGG 1045
301 sp.....ArgGlyLysMetArgAspLeuPhe...ThrPro 311
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1046 ATGTACCGAAAAAGCTGAGCCCTTCATTTGACAGACCTGTTCCCGACGCG 1095
312 HisPheArgTrpThrThrLeuLeuLeuTrpPheIleTrpPheSerAsnAl 328
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1096 CGCCTGAGGAAGCGACCTTCATCTGATGATGATGATGATGATGATGATG 1145
328 apheSerTyrGlyLeuValLeuLeuThrThrGluLeuPheGlnAlaG 345
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1146 TGTGCTCTATCAGGGGCTCATCTCTG.....CACATGG 1177
345 lYAspValCysGlyIleSerSerArgLysLysAlaValGluAlaLysCys 361
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1178 GCGCCACCGCGGG..... 1191
362 SerLeuAlaCysGluTyrLeuSerGluGluAspTyrMetAspLeuLeuTr 378
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1192 .....AACCTCTACCTGATTTCCCTTTA 1214
378 pThrThrLeuSerGluPheProGlyValLeuValThrLeuTrpIleIleA 395
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1215 CTCCTGCTCTGTCGAAATCCCGGGGCTTCATAGCCCTCATCACCATTG 1264
395 spArgLeuGlyArgLysLysThrMetAlaLeu..... 405
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1265 ACCGCTGGCGCGCATCTACCCCATGCGCTGTCAAATTTGTTGGCGGG 1314
406 .....CysPheValIlePheSerPheCysSerLeu.....Le 416
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1315 GCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1364
416 uLeuPheIleCysValGlyArgAsnValLeuThrLeuLeuPheIleA 433
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1365 CATAATCATGTGTGGCGGAATGGAATCACCATTGCAATA..... 1407
433 laArgAlaPheIleSerGlyPheGlnAlaAlaTyrValTyrThrPro 449
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1408 .....CAAATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1431
450 GluValTyrProThrAlaThrArgAlaLeuGlyLeuGlyThrCysSerGl 466
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1432 GAGCTGATCCACATTCGTCAGGAACCTCAGATGATGATGATGATGATGAT 1481
466 yMetAlaArgValGlyAlaLeuIleThrProPheIleAlaGlnValMetL 483
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1482 CCTGTGTGACATAGGTGGGATATACACCCCTTCATAGTCTTCAGCGCTGA 1531
483 euGluSerSerValTyrLeuAlaValTyrSerGlyCysCysLeu 499

```

```

1532 GGGAGGTCCTGGCAGGCTTGCCTCATTTGTTCGGTGTTGCCCGK 1581
||||| ||| ||| :||| ||| :||| |||
500 LeuAlaIleuAlaSerCysPheLeuProIleGluThrLysGlyGI 516
||||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1592 CTTGCCCGGGAGTACGCTACTTCTCCA...GACACCAAGGGGACGC 1628
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
516 yLeuGlnGluSer 520
|||| ||||::
1629 TTTGCCAGACC 1641
|||| ||||::

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-040-444-5

seq_documentation_block:
; Sequence 5, Application US/09040444
; Patent No. 6063766
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein Which Effects The
; TITLE OF INVENTION: Transport of cationic xenobiotics and/or Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan,Henderson,Farabow,Garrett & Dunner, L.L.P.
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,444
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1885 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-040-444-5

alignment_scores:
Quality: 425.50 Length: 421
Ratio: 1.766 Gaps: 11
Percent Similarity: 57.245 Percent Identity: 30.166

alignment_block:
US-09-911-667A-2 x US-09-040-444-5 ..

Align seg 1/1 to: US-09-040-444-5 from: 1 to: 1885

119 SerTrpGlnValAlaLeuLeuThrSerValPheValGlyMetMetSe 135
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
505 TCCTGGAGCTGGACCTCTTTCAGTCTGTTGAATCGGGGCTTCTTCTT 554
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
135 rSerSerThrLeuTrpGlnYAsnIleSerAspGlnTyrGlyArgLysThrG 152
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


223 PheGluValValLeuAlaValPheValMetProSerLeuGly..... 236
786TTCAATTCCTGGCCTGGCCATATGCGGT 814
237TrpArgTrpLeuLeuLeuSerAlaValProLeuLeuLeu 251
815 TCCTCAGTGGCGCTGGCTACAGTTGTCGGTCTGCTGCTCTCTCATCT 864
251 heAlaValCysPheTrpLeuProGluSerAlaAlaArgTyrAspValLeu 267
865 TCTCCTTGTGTCTGGTGGGTACAGAGTCCATACGCTGGCTGGTCTG 914
268 SerGlyAsnGluGluLysAlaAlaLeaThrLeuLysArgIleAlaThr... 283
915 TCTGAAAATCTTCACGAGCTCTGAAGACACTCAACAGCTGGGTACCT 964
284GluAsnGlyAlaProMetProLeuGlyLysLeuIleI 296
965 CAACGGCAAGAGGAGGAGGAAAGCTCAGTGTGGAGGAGCTGAAGT 1014
296 leSerArgGlnGluAsp.....ArgGlyLysMetArg 306
1015 TCACCTTCGACGAGACATCACCCTCAGCCAAGGTCAAATATGGCTTATCT 1064
307 AspLeuPhe...ThrProHisPheArgTrpThrThrLeuLeuLeuTrp 322
1065 GACTTGTTCGAGTGTCCATCTCGGCGGTGACCTTCTCTCTCTCTCT 1114
322 etLeTrpPheSerAsnAlaPheSerTyrTyrGlyLeuValLeuLeuThr 339
1115 GGCCTGGTGTGCTACTGGCTTTGGCTACTACAGTTTGGCTATGGGATG 1164
339 hrCLeuLeuPheClnAlaGlyAspValCysGlyIleSerSerArgLysLys 355
1165 AGAATTT.....GGAGTCAACATC..... 1184
356 AlaValGluAlaLysCysSerLeuAlaCysGluTyrLeuSerGluGluAs 372
1185TACATA..... 1190
372 pTyrMetAspLeuLeuTrpThrThrLeuSerGluPheProGlyValLeu 389
1191CTCCAGATCATCTTCGGTGGGGTT...GACATATCCGCCCAAGTCA 1233
389 alThrLeuTrpIleIleAspArgLeuGlyArgLysLysThrMetAlaLeu 405
1234 TCACAACTCTCCATAGTTATCTGGCCGGCGCATCACTCAGGGCTTC 1283
406 CysPheValIlePheSerPheCysSerLeu...LeuLeuPheIleCysVa 421
1284 CTCTCATCTCTGGCAGAGTGGCCATCTCGCCCTCATCTTGTGTCTTC 1333
421 lGlyArgAsnValLeu...ThrLeuLeuPheIleAlaArgAlaPheI 437
1334 AGAATTCAGCTCTTGACACAGCACTGGCTGTATTTGGGAGGAGTGC 1383
437 leSerGlyGlyPheGlnAlaAlaTyrValTyrThrProGluValTyrPro 453
1384 TGTCTGGCTCTTACAGTCGCCTCTTCCTCTACACAAGTACAGCTACGCT 1433
454 ThrAlaThrArgAlaLeuGlyLeuGlyThrCysSerGlyMetAlaArgVa 470
1434 ACAGTCTCAGGCAACAGGTATGGGTATCAGTAACTATATGGCTCGAGT 1483
470 lGlyAlaLeuIleThrProPheIle.....AlaGlnValM 482
1484 GGCAAGTATGATAGCCCACTGGTGAAATCACCGGAGAACTGCAGCCCT 1533
482 etLeuGluSerSerValTyrLeuThrLeuAlaValTyrSerGlyCysCys 498
1534 TCATCCCTAAATGTCATCTTTGGACATGACTCTACTGGGAGGC..... 1577
499 LeuLeuAlaAlaLeuAlaSerCysPheLeuProIleGluThrLysGlyVl 515

```

1578 .....AGTCTGCTTCCTTTCTGTGGACACCTCAATCG 1612
      ||| ||||| ::: :||||||| |||::
515 yGlyLeuGlnGluSerSerHis.....ArgGluTrpGlyGlnGlu 528
      ||| ||||| :: :||||||| |||::
1613 GCCTTACCAGAAACTATCGAGGACATACAAGACTGTTACCAGCAA 1658
      ||| ||||| ::: :||||||| |||::

seq_name: /cyn2_6/ptodata/2/ina/6B_COMB.seq:us-09-103-840A-1

seq_documentation_block:
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103.840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H3RV
US-09-103-840A-1

alignment_scores:
          Quality: 230.00           Length: 476
          Ratio: 0.975              Gaps: 14
Percent Similarity: 49.580    Percent Identity: 22.479

alignment_block:
US-09-911-667A-2 x US-09-103-840A-1 ..

Align seg 1/1   to: US-09-103-840A-1   from: 1   to: 4411529

      86 LeuSerValLeuThrGlyLeuAlaTrpMetAlaAspAlaMetGluMetMe 102
         ::::||||| ||| ||| |||
3717266TGTCGGGTCTCTACGGGTACGACCTTTCCGCCATCGCGGTGCTGCT 3717315

      102 ttleLeuSerlleLeuAlaProGlnLeuHisCysGluTrpArgLeuProS 119
         : |||||
3717316GTCTCTCACGCAG.....GAATTGCACTCACCA 3717344

      119 exTrpGlnValAlaLeuLeuThrSerValValPheValGlyMetMetSer 135
         : ::: ||||| ||||| ::: ||||| |||||
3717345CTCGAGAACAGGATGCTGACCAACCACGGCGGTGCTCGCCAGATCGCC 3717394

      136 SerSerThrLeuTrpGlyAsnIleSerAspGlnTyrGlyArgLysThrGl 152
         : ::: ||| ||||| ||||| ::: ||||| |||||
3717395GGGGCGCTTGGCGCGGCATCTCGCCAACGGCATCGGACGCAAGAATTC 3717444

      152 yLeuTySlleSerValLeuTrpThrLeuTyrTyrGlyIleLeuSerAlap 169
         : ::: :||| :::: ||||| |||||
3717445GGTGGTGCTCATCGTCGGCGGTACGCAGTGTGCGCCTCTCTCGGCCGA 3717494

      169 heAlaProValTyrSerTrpIleLeuValLeuArgGlyLeuValGlyPhe 185
         : ::: :||| ||||| ||||| ::: ||||| |||||
3717495CTCGGTGCCGTACCGATGCTGTTGGTGGCGGCTGCTGCTGGGTGTG 3717544

      186 GlyIleGlyGlyValProGlnSerValThrLeuTyr...AlaGluPheLe 201
         ||||| ||||| ||| :||| |||||
3717545ACAATCGCGCTGTCGGTGTGGTGGTGGCGGTATGTCGCCGAGTCGCG 3717594

      201 uProMetLysAlaArgAlaLysCysIleLeuLeuIleGluValPheTrpA 218
         ||| :||| ||||| :||| :||| :|||
3717595GCGCGCGCGGTGCGGTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3717644

```



```

732 CTCCTCTGTGCTACCAACCGCATCGGCCAGTGGGATCTGGTGTGTA 781
119 ....SerTrpGlnValAlaLeuLeuThrSerValValPheValGlyMetM 134
    ::::|||||  |||  :::::|||||
782 CCTGGGCTGCAGGTATCTCGAGCAGATCTCTTCCATCTGGCTTG 831
134 etSerSerThrLeuTrpGlnPheValSerAspGlnTrpGlyArgLys 150
    ::::|||||  |||  ::::|||||
832 CCTCGGCTACCTGTTCTGGGTACCCGACAGACAGATTGGCCGTCG 881
151 ThrGlyLeuLysIleSerValLeuTrpThrLeuTrpTyr..... 163
    |||:::  |||||  |||||
882 ...GGGATT.....GTGCTGCTGACCTTGGGCTGGTGGCCCTG 919
164 GlyIleLeuSerAlaPheAlaProValTyrSerTrpPheLeuValLeu 180
    |||:::  |||  :::::|||||
920 TGGAGTAGGAGGGCTGCTGAGGCTCTCCACAGAGGCTATGGCCCTC 969
180 rGlyLeuValGlyPheGlyIleGlyValProGlnSerValThrLeu 196
    |||:::  |||||  |||||
970 GATTCCCTTGGGCTTCTGCTGGGCTGTGACCTGGGTGTCTACCTG 1019
197 Tyr...AlaGluPheLeuProMetLysAlaArgAlaLysCysIleLeu 212
    |||:::  |||  :::::|||||
1020 ATGCGCCTGGAGCTGCGACCCCAACAGAGGCTTCGGGTGGCCCTGC 1069
212 uIleGluValPheTrpAlaIleGlyThrValPheGluValValLeuAla 229
    |||:::  |||||  |||||
1070 AGGGGAGTTG.....GTGGGGTGGGAGGCACTTCTGTCCTGG 1110
229 alPheValMetProSerLeuGlyTrpArgTrpLeuLeuLeuSerAla 245
    ::::|||||  |||  |||||
1111 GCCTGGCCCTGCTCTAAGGATTGGGATTCCTACAGCGAATGATCACC 1160
246 ValProLeuLeuPheAlaValLeuCysPheTrpLeuPro..... 259
    ::::|||||  |||  :::::|||||
1161 GCTCCCTGCATCTCTTC.....CTGTTTATGGCTGGCTGGCTTTGTT 1204
260 ....GluSerAlaArgTrpAspValLeuSerGlyAsnGlnGluLysAla 275
    ::::|||||  |||  |||||
1205 CTGAGTCCCGCAGCGTGGTGTAGTGAAGCGCAGATTGAGGAGCTC 1254
275 leAlaThrLeuLysArgIleAlaThrGluAsnGlyAlaProMetProLeu 291
    ::::|||||  |||  |||||
1255 AGTCTGTGCTGAGGATCTGCTGAGCGAAAC.....CGGCCCAT 1295
292 GlyLysLeuIleIleSerArgGlnGluAspArgGlyLysMetArgAspLe 308
    |||:::  |||||  |||||
1296 GGGCAGATGCTG.....GGGAGGAGGCCAGGAGGCCCTGCGAGGACCT 1339
308 u.....PheThrProHisPheArgTrpThrL 318
    |||  |||  |||
1340 GGAGAATACCTGCCCTCTCCCTGCAACATCTCTCTTTCTTTGTTGCC 1389
318 euLeu.....LeuTrp.....PheIleTrpPheSer 326
    |||||  ::::|||||  ::::|||||
1390 TCCTCACTACCGCAACATCTGGAATAATCTGTTATCTCGGCTTCACC 1439
327 AsnAlaPheSerTyrGlyLeuValLeuLeuThrThrGluLeuPheGln 343
    |||  |||  ::::|||||
1440 AAC...TTTATGCCCATTGCACTTCGCCACTGCTACCGAGCTGTGGAGG 1486
343 nAlaGlyAsp.....ValCysGlyIleSerSerArgLysL 355
    ::::|||||  ::::|||||  ::::|||||
1487 AGGAGGAGGCCCATCGGACTTCTACCTGTGCTCTGCTGGCCAGCGGCA 1536
355 ysAlaValGluAlaLysCysSerLeuAlaCysGluTyrLeuSerGluGlu 371
    |||  |||  ::::|||||
1537 CCGCA.....GCCCTGGCTGTCTTCCTG..... 1562
372 AspTyrMetAspLeuLeuTrpThrLeuSerGluPheProGlyValLe 388
    |||||
1563 .....GGGTCAC 1570

```

```

388 uValThrLeuTrpIleIleAspArgLeuGlyArgLysLysThrMetAlaL 405
    |||  |||||  ::::|||||
1571 CGTG.....GACCGATTGGCGCGCGGCGCATCTTCTCTTC 1605
405 euCysPheValIlePheSerPheCysSerLeuLeuLeuPheIleCysVal 421
    |||:::  |||||  |||||
1606 TCTCATGACCTTACCGGATTCCTCCCTGGTCTCTG..... 1643
422 GlyArgAsnValLeuThrLeuLeuPheIleAlaArgAlaPheIleSe 438
    |||  |||  ::::|||||
1644 .....CTGGGCTCTGGGATTATCTGAACGAGGCTGCCATCAC 1681
438 r.....GlyGlyPhe.....GlnAlaAlaTyrValTyrT 448
    :  :  |||  |||||
1682 CACTTCTCTGCTGCTGGCTCTTCTCTCCCAAGCTGCCGCGCATCTCA 1731
448 rr.....ProGluValTyrProThrAlaThrArgAlaLeuGly 460
    |||||  |||||  ::::|||||
1732 GCACCTCTTCTGCTGCTGAGTCACTCCACACACTGTCGGGCGCGTGC 1781
461 LeuGlyThrCysSerGlyMetAlaArgValGlyAlaLeuIleThrProPh 477
    |||||  ::::|||||  |||
1782 CTGGCCTGATCATGGCTCTAGGGCGCTTGGAGACTGAGCGCCG... 1829
477 eIleAlaGlnValMetLeuGluSerSerValTyrLeuThrLeuAlaValT 494
    ::::|||||  ::::|||||  ::::|||||
1830 .GCCAGCGCTCCACATGGGCGCATGGAGCTTCTGTCAGCACGCTGGTGC 1878
494 yrSerGlyCysCysLeuLeuAlaAlaLeuAlaSerCysPheLeuProIle 510
    ::::|||||  ::::|||||  ::::|||||
1879 TGGCGGCTGCGCCCTCTCTGTCATCTCAGCATTCATGCTGCTGCCG... 1925
511 GluThrLys 513
    |||||
1926 GAGACCAAG 1934

```

seq_name: /cgn2 6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2

seq_documentation_block:

```

; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

```

alignment_scores:
    Quality: 246.50      Length: 534
    Ratio: 0.970        Gaps: 19
    Percent Similarity: 47.566    Percent Identity: 22.097

alignment_block:
US-09-911-667A-2 x US-09-103-840A-2/rev ..

```


Align seg 1/1 to reverse of: US-09-103-840A-2 from: 1 to: 4403765

12 ProValValLys.....PheArgArgThrGlyGluSerAlaArgSe 25
|||||.....
2147670CCCAATCGAATCCGGTCACTCGCGCGCATCAGAAAGCCACCAATT 2147621

25 rGluAspAspThrAlaSerGlyGluHisGluValGlnIleGluGlyValH 42
|||||.....
2147620GCCCGCATCTGCCACGGAC..... 2147602

42 isValGlyLeuGluAlaValGluLeuAspGlyAlaAlaValProLys 58
|||||.....
2147601...ATGCGACCGCGCTGTTCGCCATCTCCCGGACTCGCGTGGCGGC 2147554

59 GluPheAlaAsnProThrAspAspThrPheMetValGluAlaValG 75
::: |||||
2147553CCGGTACGCACGATGCACAG.....GTGGC 2147528

75 uAlaIleGlyPheGlyLysPheGlnTrpLysLeuSerValLeuThrGlyL 92
|||||.....
2147527AGCACCGCGTTGACCGGCACCGAAGCGCTTTATGGCATCTTTCT 2147478

92 euAlaTrpMetAlaAspAlaMetGluMetMetIleLeuSerIleLeuAla 108
|||||.....
2147477TGGTTGGACGATGGATCGTTGATTAATTCTCGTGGTGGTCTAT 2147428

109 ProGluLeuHisCysGluTrpArgLeuProSerTrpGlnValAlaLeuLe 125
::: |||||
2147427GCCGACATGCCAACGACCTTCCACACACCAAGACCGCATCGCTTCT 2147378

125 uThrSerValValPheValGlyMetMetSerSerSerThrLeuTrpGlyA 142
|||||.....
2147377ACACCGCCACCTGGCCATGGCCCGCGGTGGGTGGTGGTGGTGGG 2147328

142 snIleSerAspGlnTyArgLysThrGlyLeuLysIleSerValLeu 158
|||||.....
2147327TATGGCGGACCGGTGGCGCGCGGTCCGCTGATGGTGGACGTGTCG 2147278

159 TrpThrLeuTyTrpGlyIleLeuSerAlaPheAlaProValTySerTr 175
|||||.....
2147277TTCATTCCGTGATGTTCGTGCGCATTCGCACCACTTCACCGT 2147228

175 pIleLeuValLeuArgGlyLeuValGlyPheGlyIleGlyGlyVal...P 191
::: |||||
2147227GCTGGTATCTCGACTGCTCTACGGCATTTGCATGGCGGAGTGGG 2147178

191 roGlnSerValThrLeuTyArgAlaGluPheLeuProMetLysAlaArgAla 207
::: |||||
2147177GGCTGGGTGCGCTGTCCATGAGAAAGGTCCAGCCGAGCGCGTGGC 2147128

208 LysCysIleLeuLeuIleGluValPheTrpAlaIleGlyThrValPheG 224
|||||.....
2147127CTCTCCGGCTGCTGCAGAGGGCTACGCGTTCGGCTATCTGCTGGC 2147078

224 uValValLeuAlaValPheValMetProSerLeuGly.....TrpArgT 239
|||||.....
2147077CAGTGTGGCAGCGCTGGTGGTGAATGAACCTGGCTAGGCGTCTCGTGGCGGT 2147028

239 rpLeuLeuIleLeuSerAlaValProLeuLeuLeuPheAlaValLeuCys 255
|||||.....
2147027GTTTTCGGCTTAAGCATATCCCGGCTCTGATCAGCTTGATCATCCGG 2146978

256 PheTrpLeuProGluSerAlaArgTyArgValLeuSerGlyAsnGlnG 272
::: |||||
2146977TACCGGGTAAGGAATCC.....GAAGTCTGGGAAGCGCGCAGGA 2146937

272 uLysAlaIleAlaThrLeuLysArgIleAlaThrGluAsnGlyAlaProM 289
::: |||||
2146936CCGATCGCGCTCACCAGACCCGATCCCGGATGCTCCTGGGTAAACCCC. 2146888

289 etProLeuGlyLysLeuIleIleSerArgGlnGluAspArgGlyLysMet 305
|||||.....
2146887.....GCAATTGTGGCGCG..... 2146873

306 ArgAspLeuPheThrProHisPheArgTrpThrThrLeuLeuLeuTrp... 321
|||||.....
2146872.....TTTCTACCTAGTGTCTGTGATGACCCG 2146844

322 PheIleTrpPheSerAsnAlaPheSer.....TyTrpGlyLeuVal 336
|||||.....
2146843CTTCAACTGGATGAGCCAGGCGCATCTACCGACCTTCCTGA 2146794

336 euLeuThrThrGlu..... 340
|||||.....
2146793CCGCGACTACCGCATGATGTCGGGTTTTCAGCGCTACCGCCAGGTGG 2146744

341 ...LeuPheGlnAlaGlyAspValCysGlyIleSerSerArgLy 354
::: |||||
2146743ATAGTGGTGTCTACAACTCGCGCCCATCATCGGC..... 2146708

354 sLysAlaValGluAlaLysCysSerLeuAlaCysGluTyTrpLeuSerGluG 371
2146708..... 2146708

371 luAspTyTrpMetAspLeuLeuTrpThrThrLeuSerGluPheProGlyVal 387
|||||.....
2146707.....GGTCTGGCGTTCCGACGTTGTCCGAG..... 2146681

388 LeuValThrLeuTrpIleIleAspArgLeuGlyArgLysThrMetAl 404
|||||.....
2146680.....CGTTTCAGCGCGCTTACACCATCGT 2146655

404 aLeuCysPheVal.....IlePheSerPheCysS 414
::: |||||
2146654TTTCTGCGCGCTGGACTGCGCATCGTGGCGGTGTTCGTTACTCGC 2146605

414 erLeuLeuLeuPheIleCysValGlyArgAsnValLeuThrLeuLeu 430
::: |||||
2146604GCATCGCGCGATGTTGTGCTGGGTTCG..... 2146576

431 PheIleAlaArgAlaPheIleSerGlyGlyPheGlnAlaAlaTyValTy 447
|||||.....
2146575T'TTTTCATCAAGTCTTCGTACAGGTGCTGGGGTGTATCCCGCGCA 2146526

447 rThrProGluValTyTrpThrAlaThrArgAlaLeuGlyLeuGlyThrC 464
|||||.....
2146525TCTGACCGAGATGTCGCGACGCTATCCGG.....GGCGTCT 2146488

464 ySerGlyMetAla...ArgValGlyAlaLeuIleThrProphe..... 477
|||||.....
2146487ACCCCGCGGTAACTACCGATCGTAATCTGCTGGCGGTCAACCTT. 2146438

478 ...IleAlaGlnValMetLeuGluSerSerValTyTrpLeuThrLeuAlaVa 493
|||||.....
2146437CCTATTCAAGAGCGTCTGGCCGAATCCCATGGCTAC..... 2146402

493 lTyTrpSerGlyCysLeuLeuAlaAlaLeuAlaSerCysPheLeuProI 510
|||||.....
2146401.....CCCTTCGCTTGGCGCGACGATCGTGGCGG 2146371

510 le 510
:::
2146370TG 2146369

seq_name: /cqn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-1

seq_documentation_block:

; Sequence 1, Application US/09103840A
; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; TITLE OF INVENTION: TUBERCULOSIS

```

; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103.840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
; US-09-103-840A-1

```

```

alignment_scores:
  Quality: 246.50      Length: 534
  Ratio: 0.970        Gaps: 19
  Percent Similarity: 47.566  Percent Identity: 22.097

```

```
alignment_block:
```

```
US-09-911-667A-2 x US-09-103-840A-1/rev ..
```

```
Align seg 1/1 to reverse of: US-09-103-840A-1 from: 1 to: 4411529
```

```

12 proValVallys.....PheArgThrGlyGluSerAlaArgse 25
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2150410CCCCAATATGCGATCCGGTCACTCGCGCGGCATCAGAAAGGCACCAT 2150361
25 rGluAspThrAlaSerGlyGluHisGluValGlnIleGluGlyValH 42
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
2150360GCCCGCATCTGCCACGGAC..... 2150342
42 isValGlyLeuGluAlaValGluLeuAspGlyAlaAlaValProLys 58
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:
2150341...ATGCGAGCGCGCGTGTTCGCATGTCGCCGACTCGCGGTGCCGCG 2150294
59 GluPheAlaAsnProThrAspAspThrPheMetValGluAspAlaValG 75
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:
2150293CCGTACGCCACGATGCAGAC.....GTGCG 2150268
75 uAlaIleGlyPheGlyLysPheGlnTrpLysLeuSerValLeuThrGlyL 92
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
2150267ASCACCGCGGTGACCGCGGACCAAGCGTCTTTATGGCATCTTTCT 2150218
92 euAlaTrpMetAlaAspAlaMetGluMetIleLeuSerIleLeuAla 108
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2150217TGGGTGGACGATGCGTTCGATTCTCTCGTGGTCTAT 2150168
109 ProGlnLeuHisCysGluTrpArgLeuProSerTrpGlnValAlaLeuLe 125
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:
2150167GCCGACATCGACAGACCTTCCACACCAAGACCGACGTCGCTTTCT 2150118
125 uThrSerValValPheValGlyMetMetSerSerThrLeuTrpGlyA 142
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2150117CACCACCGCCACCTGGCCATGCGCCCGTGGGTGGCTGCTGTCGGGC 2150068
142 snIleSerAspGlnTyrclyArgLysThrGlyLeuLysIleSerValLeu 158
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:
2150067TATGGCGCGACCGGTGGCGCGCGGTGCGCTGATGTCGACGTGTCG 2150018
159 TrpThrLeuTyrclyGlyLeuSerAlaPheAlaProValTySerTr 175
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:
2150017TTCTATTCGGTATCGATTCTTCGCGCATTCGACCCCACTTCCACCG 2149968
175 pIleLeuValLeuArgGlyLeuValGlyPheGlyIleGlyGlyVal...P 191
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:
2149967GCTGCTGATCTCGACTGCTCTACGGCATTTGGCATGGCGGAGATGGG 2149918
191 roGlnSerValThrLeuTyrclyAlaGluPheLeuProMetLysAlaArgAla 207
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:
2149917GCTGGTGGCGCGCTGTCCATGGAGAGGTTCAGCGCGACGCGCGTGGC 2149868
208 LysCysIleLeuLeuGluValPheTrpAlaIleGlyThrValPheG 224
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

2149867GTCTTCTCCGGGCTGTGTGACGAGGGCTACGGTTCGGCTATCTGTGGC 2149818
224 uValValLeuAlaValPheValMetProSerLeuGly.....TrpArgT 239
||| :|||:|||||:|||||:|||||:|||||:|||||:
2149817CAGTGTGGCAGCGCTGTGTGATGAACCTAGGCTGTGTGGCGGT 2149768
239 rPleuLeuIleLeuSerAlaValProLeuLeuLeuPheAlaValLeuCys 255
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2149767GTGTTCGGCTTAAGCATCATCCGCTCTGATCAGCTTCATCATCCGG 2149718
256 PheTrpLeuProGluSerAlaArgTyrclyAspValLeuSerGlyAsnGlnI 272
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:
2149717TACCGGTAAAGGATCC.....GAAGTCTGGGAGCGCGCAGGA 2149677
272 uLysAlaIleAlaThrLeuLysArgIleAlaThrGluAsnGlyAlaProM 289
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:
2149676CCGCTCGCTCACCAGACCGCAATCCGCGATGCTCTGGTAAACCCC. 2149628
289 etProLeuGlyLysLeuIleSerArgGlnGluAspArgGlyLysMet 305
|||:|||||:|||||:|||||:|||||:|||||:|||||:
2149627.....GCAATTGTGGCGCGG..... 2149613
306 ArgAspLeuPheThrProHisPheArgTrpThrThrLeuLeuLeuTrp... 321
||| :|||:|||||:|||||:|||||:|||||:|||||:
2149612.....TTGTCTACCTAGTGTGTGTGATGACCGC 2149584
322 PheIleTrpPheSerAsnAlaPheSer.....TyrclyLeuValL 336
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2149583CTTCAACTGGATGACCGACCGCATCTACCCGACCTTCCTGA 2149534
336 euLeuThrThrGlu..... 340
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2149533CCGCGACTACGGATCATGTCGCCGTTTGTCCAGCTCACCGCCAGGTGG 2149484
341 .....LeuPheGlnAlaGlyAspValCysGlyIleSerSerArgLy 354
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:
2149483ATAGTGTGTGATCTACAACATCGCGCGCATCATCGGC..... 2149448
354 sLysAlaValGluAlaLysCysSerLeuAlaCysGluTyrclyLeuSerGluG 371
2149448..... 2149448
371 luAspTyrclyMetAspLeuLeuTrpThrLeuSerGluPheProGlyVal 387
||| :|||:|||||:|||||:|||||:|||||:|||||:
2149447.....GGTCTGGCGTTCGGCAGCTGTGCCAG..... 2149421
388 LeuValThrLeuTrpIleIleAspArgLeuGlyArgLysLysThrMetAl 404
|||:|||||:|||||:|||||:|||||:|||||:|||||:
2149420.....CGGTTCAGCGCGCGCTTACACCATCGT 2149395
404 aLeuCysPheVal.....IlePheSerPheCys 414
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:
2149394TTTCTGTCCGCGCTGGGACTGCCGATGTCGCGCTGTCGTTACTCGC 2149345
414 erLeuLeuLeuPheIleCysValGlyArgAsnValLeuThrLeuLeuLeu 430
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:
2149344GCATCGCGCGATGTGTGCTGGTTCG..... 2149316
431 PheIleAlaArgAlaPheIleSerGlyGlyPheGlnAlaAlaTyrclyVal 447
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2149315TTTTTGTGCAAGTCTCTGACAGGGTCCCTGGGTGTGTGATCCCGCGCA 2149266
447 rThrProGluValTyrclyProThrAlaThrArgAlaLeuGlyLeuGlyThrC 464
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:
2149265TCTCAGCGAGATGTCCGACGACGCTATCCGG.....GCGTCT 2149228
464 ysSerGlyMetAla...ArgValGlyAlaLeuIleThrProPhe..... 477
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2149227ACCCCGCGGTAACTTACGCTGATCTGCTGCGCGCGGTTCACCTT 2149178
478 ...IleAlaGlnValMetLeuGluSerSerValTyrclyLeuAlaVala 493
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
2149177CCTATTCAAGAGCGTCTCGCGCAATCCCATGGCTAC..... 2149142

```

826 GCCTGGGAGCGCTCAACCAACTGGCCATTGTTATCGCATTCGTCTGATCC 875

232 tProSerLeuGly.....TtpArgTf 239
|||||
876 CCAGGTGCTGGCTTGGAGTCCCTCCCTGGGCACTGCCACGCTGTGGCCAC 925

239 rPleuLeuIleLeuSerAlaValProLeuLeuLeuPheAlaValLeuCys 255
|||||
926 TGTCTCTGGGCCCTACAGTAGTACTGCCCTCTGCTGCACTGGCTGCTCTG 975

256 PheTrpLeuProGluSerAlaAArgTytrAspValLeuSerGlyAsnGlnG 272
:::
976 CCCTTCTGTCCTCCGAGAGCCCGCTACCTCTACATCATCCACAATCTCGA 1025

272 u.....LysAlaIleAlaThrLeuLysArgIleAla..... 282
| ||| ::|||
1026 GGGGCTGCCAGAAAGAGTCTGAAGCGCTGACAGGCTGGGCGGATGTTT 1075

283ThrGluAsnGly 286
|||||

1076 CTGGAGTGTGGCTGAGCTGAAGATGAGAGCGGAAGCTGGAGGCTGAG 1125

287 AlaProMetProLeuGlyLysLeuIleIleSerArgGlnGluAspArgG 303
||||: ||| ::|||: ||| ||| ::
1126 CGGCCACTGTGCTGCTCCAGCTCTCTGGCAGCGCTACCCAC..... 1167

303 yLysMetArgAspLeuPheThrProHisPheArgTrpThrLeuLeuL 320
:::
1168CGGCGCCCTGATCATTTG 1186

320 enuTrpPheIleTrpPheSer.....AsnAlaPheSer 330
::: ::|||
1187 CGGTGCTGTGCACTGAGCCAGCACTCTGGCATCAATGCTGTTTC 1236

331 TyrTytrGlyLeuValLeuLeuThrThrGluLeuPheGlnAlaGlyAspVa 347
|||||
1237 TATTAT.....TCGACCAGCATCTTCGACAGCAGCGGCT 1271

347 lCysGlyIleSerSerArgLysLysAlaValGluAlaLysCysSerLeuA 364
:::
1272 A.....GGCAGCGCTGCTATGCCACCATAG 1297

364 lAcysGluTytrLeuSerGluGluAspTytrMetAspLeuLeuTrpThrThr 380
:::
1298 GAGCT.....GGTGTGGTCAACACAGTCTTCACC... 1326

381 LeuSerGluPheProGlyValLeuValThrLeuTrpIleIleAspArgLe 397
|||||
1327TTGGTCTCGGTGTGTTGGTGAGCGGGC 1355

397 uGlyArgLysLysThr.....MetAlaLeuCysP 407
|||||
1356 GGGGCGCGGACGCTCCATCTCTCTGGGCTGGCGGCATGTGTGGCTGTG 1405

407 heValIlePheSerPheCysSerLeuLeu..... 417

1406 CCATCTGTATGACTGTGGCTCTGCTCTGGAGCGAGTCCAGCCCATG 1455

418 ...PheIleCysValGlyArgAsnValLeuThrLeuLeuPheIleAl 433
:::
1456 AGCTAGTCT.....TCCATTGTGGCCATCTTTGGCTCTGGTGGC 1493

433 aArgAlaPheIleSerGlyGly.....PheGlnAlaAlaTytrV 446
| |||: |||
1494 ATTTTGTGAGATTGGCCCTGGCCCATTCCTTTGGTTTCATCTGGCGCGAG 1543

446 alTytrProGluValTytrProThrAlaThrArgAlaLeuLeuGly 462

1544 TCTTCAGCAGGAGCCCGCGGAGCCATGGCTGGTGTCTCTCC 1593

463 ThrCysSerGlyMetAlaArgValGlyAlaLeuIleThrProPheIleAl 479
::: ::|||
1594 AACTGGCAGGCAACTTCATCATTTGGC...ATGGTTTCCAGTATGTTGC 1640


```
303 .....GlyLysMetArgAspLeuPheThrProHisPheArgTrpThrT 317
||||| ..... |||||
19730 GGGCTGGGATCGCGCGTTCGCTG.....A 19705
317 hrLeuLeuLeuTrpPheLeTrpPheSerAsnAlaPheSerTyrTyrGly 333
||||| ..... |||||
19704 CGCTGTC.....GCCGTCCTCGTGTGATCGGC 19679
334 LeuValLeuLeuThrThrGluLeuPheGlnAlaGlyAspValCysGlyI 350
||||| ..... |||||
19678 CTTGGTCTGCTTCGCTG.....CT 19662
350 eSerSerArgLysAlaValGluAlaLysCysSerLeuAlaCysGluT 367
||||| ..... |||||
19661 GCCCAAGCGCGCGCGCTGGAGCG..... 19634
367 yrLeuSerGluGluAspTyrMetAspLeuLeuTrpThrThrLeuSerGlu 383
19634 ..... 19634
384 PheProGlyValLeuValThrLeuTrpIleLeuAspArgLeuGlyArgly 400
||||| ..... |||||
19633 .....ATGGTCCCGTACGGCGTGTTCGGCGCGGTGGCATCAC 19596
400 slsYThrMetAlaLeuCysPheValIle.....P 410
||||| ..... |||||
19595 GATGACACCGCGCTCAACTTCCACATCGGCGTCCGATCTTCGGTACGG 19546
410 heSerPheCysSerLeuLeuLeuPheIleCysValGlyArgAsnVal... 425
||||| ..... |||||
19545 TCTCCACCTGCGCTGTTCCTGCGAGCTGGTGCAGGGCGCGAGTCCACC 19496
426 .....LeuThrLeuLeuLeuPheIleAlaArgAlaPheIleSerG 439
||||| ..... |||||
19495 GTCCGCGGACTGCTGATCATCCGCTGATGACCGGGCGGATGCTCTCCA 19446
439 yGlyPheGlnAlaAlaTyrValTyrThrProGluValTyrProThrAlaT 456
||||| ..... |||||
19445 GACGATCTGCGCAAGATCAAGAACTGGAACCGCTACAAGAAGCGCG 19396
456 hrArgAlaLeuGlyLeuGlyThrCysSerGly.....MetAla 468
||||| ..... |||||
19395 CG....ATCGTGGCGCTCGGCTCGATGGCGGAGACACTGCTGCTCTCC 19349
469 ArgValGlyAlaLeuIleThrProPheIleAlaGlnValMetLeuGlu... 484
||||| ..... |||||
19348 GCCCGCGCGCG...GACACCCCGCTCGCGGTGATCGTGTGATCGCGGC 19302
485 .....SerSerValTyrLeuThrLeuAlaVal 493
||||| ..... |||||
19301 CTGGGTGGGCTTCGGCATCGGCGCTCTCCAGACCGCTACACCCCTGGCGA 19252
493 alTyrSer 495
|||||
19251 TCCAGAGC 19244
```

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-474-933-1

seq_documentation_block:

; Sequence 1, Application US/08474933

; Patent No. 5866410

; GENERAL INFORMATION:

; APPLICANT: Ryan, Michael J.

; APPLICANT: Lotvin, Jason A.

; APPLICANT: Strathy, Nancy

; APPLICANT: Fantini, Susan E.

; TITLE OF INVENTION: Cloning of the biosynthetic pathway for

; TITLE OF INVENTION: chlortetracycline and tetracycline formation and cosmids

; TITLE OF INVENTION: useful therein

; NUMBER OF SEQUENCES: 1

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: American Cyanamid Company

; STREET: One Cyanamid Plaza

```
CITY: Wayne
STATE: New Jersey
COUNTRY: USA
ZIP: 07470
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,933
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/125,468
FILING DATE: 22-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Tsevdos, Estelle J
REGISTRATION NUMBER: 31,145
REFERENCE/DOCKET NUMBER: 31,255-02
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201)831-3241
TELEFAX: (201)831-3305
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 30001 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-474-933-1
```

alignment_scores:
Quality: 151.50 Length: 536
Ratio: 0.613 Gaps: 27
Percent Similarity: 46.082 Percent Identity: 22.575

alignment_block:

US-09-911-667A-2 x US-08-474-933-1/rev ..

Align seg 1/1 to reverse of: US-08-474-933-1 from: 1 to: 30001

```
18 ArgThrGlyGluSerAlaArgSerGluAspAspThrAlaSerGlyGluH 34
||||| ..... |||||
20560 CGGGCGGGATTCGGCAAGAACT...TGTCATTCCGCCACCGTAGCCCA 20514
34 sGluValGlnIleGluGlyVal.....HisValGlyL 45
||||| ..... |||||
20513 CTACGCTCAGTTCGACAGACCTTGTGGAATTCGGTGGGAGAAATGGAA 20464
45 euGluAlaValGluLeuAspAspGlyAlaAlaValProLysGluPheAla 61
||||| ..... |||||
20463 TGGCGAAGCCACCTCGCAGACCGCGGAGCGCTCGCGGAGAGCGCGC 20414
62 AsnProThrAspAspThrPheMetValGluAspAlaValGluAlaIleG 78
||||| ..... |||||
20413 GGCCCGCGC.....GG 20403
78 yPheGlyLysPheGlnTrpLysLeuSerValLeuThrGly.....LeuA 93
||||| ..... |||||
20402 CTTCACCTACCGCGCAG...ATCATCCCGCGCTGTCGGGCTCCTGTGG 20356
93 laTrpMetAlaAspAlaMetGluMetIleLeuSer..... 105
||||| ..... |||||
20355 CCGTGTGCTCGCGCACCTCGACACGACGATCGTCTCCACCGCGCTGGT 20306
106 IleLeuAlaProGlnLeuIleCysGluTrpArgLeuProSerTrpGlnVa 122
||||| ..... |||||
20305 ACGATCGGCGACCACTCGACGCGCGCAG...ACCGTCCAGGCGCTG... 20264
122 laLeuLeuThrSerValValPheValGlyMetSerSerThrL 139
||||| ..... |||||
```

```

410 heSerPheCysSerLeuLeuLeuPhehelleCysValIcYArgAsnVal... 425
    |||
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
19545 TCtCCACCCCTGCCGCTGTTCCTGCAGCTGGTGCAGGGGGCAGTGCACC 19496
    426 .....LeuThrLeuLeuPheIleAlaArgAlaPheIleSerGI 439
    |||:||||:||||:||||:||||:||||:||||:||||:||||:
19495 GTCGCGGAGCTGGTGATCATCCGGTGATGACCGGGGGCGATCGTCTCCCA 19446
    439 yGIlyPheGlnAlaAlaValThrThrProGluValTyrProThrAlar 456
    :: ||| ::| ::| |||
19445 GACGATCTGCCCAAGATCATCAAGAAGTGGRAACCGCTACAAGAAGCCCG 19396
    456 hrArgAlaLeuGlyLeuGlyThrCysSerGly.....MetAla 468
    :: |||:||||:||||:||||:||||:||||:||||:||||:||||:
19395 CG...ATCGTGGGCTCGGCTCGATGGCGGAGCACTGCTGCGCTCTCC 19349
    469 ArgValCylAlaLeuIleThrProPheIleAlaGlnValMetLeuGlu... 484
    :: |||:||||:||||:||||:||||:||||:||||:||||:||||:
19348 GCCGCGGGCGCG...GACACCCCGCTCGCGTGATCGTGGTGATCGCGGC 19302
    485 .....SerSerValTyrLeuThrLeuAlav 493
    |||:||||:||||:||||:||||:||||:||||:||||:||||:
19301 CTGGCTGGGCTTCGGCATCGGGCTCTCCAGACCGTCATCACCTCGCGA 19252
    493 alTyrSer 495
    :: |||
19251 TCCAGAGC 19244

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-928-692-9

seq_documentation_block:
; Sequence 9, Application US/08928692
; Patent No. 5958727
; GENERAL INFORMATION:
; APPLICANT: Brody, Howard
; APPLICANT: Iaver, Deborah S.
; APPLICANT: Lamsa, Michael
; APPLICANT: Hansen, Kim
; TITLE OF INVENTION: Methods for Modifying the Production of
; TITLE OF INVENTION: a polypeptide
; NUMBER OF SEQUENCES: 80
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5938727o No. 5958727disk of No. 5958727th #
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,692
; FILING DATE: 12-SEPT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4944.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3000 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-928-692-9

```

1529 TCCTCTTTACAGACCATCTCTCCACCACTCGACCCCTG.....TTGTC 1572

351 rSerArgIysLysAlaValGluAlaLysCysSerLeuAlaCysGluTyrL 368

1573 GGTGATCATCTCGGCTATCAATCTTGTAATCACTCTGCCTGC..... 1615

368 euSerGluGluAspTyrMetAspLeuLeuTrrThrThrLeuSerGluPhe 384

1616TCACCATA 1624

385 ProGlyValLeuValThrLeuTrpIleIleAspArgLeuGlyArgLysL 401

1625 CCT.....CATAGATTGGTAGACGCTC 1647

401 sThrMetAlaLeuCysPheValIlePheSerPheCysSerLeuLeuLeuP 418

1648 CTGCGCTGTTCTAAGTATACAGCGCATGGGTCTTAATTCGCTCTACTGG 1697

418 heIleCysValGlyArgAsnVal.....LeuThrLeuLeu 429

1698 CGCTAGCCATCTACTTCAACCTGAAGCGCTTATCCGCCATAGCAGTTCTA 1747

430 LeuPheIleAlaArgAlaPheIleSerGlyGlyPheGlnAlaAlaTyrVa 446

1748 CTTTTCGTTGCT...TCCTTCGCCCGCGGTCTAGGCCCATGCCCTTCAT 1794

446 TyrThrProGlu...ValTyrProThrAlaThrArgAlaLeuGlyLeuG 462

1795 TTTAGCCTCTCAACTCGTTGGCCGGAGCGTGC..... 1828

462 TyrThrCysSerGlyMetAlaArgValGlyAlaLeuIleThrProPheIle 478

1829GGCGCCACAGAGCTGGCGCTGGGAGCGAAGCTGGATT 1867

479 AlaGlnValMetLeu..... 483

1868 GCCACGTTTCATCGTGGCAACATTTTTCGATGTTAAACGATTGTGTGGG 1917

484GluSerSerValTyrLeuThrLeuAlaValTyrSerGlyCysL 499

1918 CGGACGAGGCAAGATCTACTGGATCTTTGGACGATGGCC.....TGTC 1961

499 euLeuAlaAlaLeuAlaSerCysPheLeuProIleGluThrLysLysGly 515

1962 TCCTCGGAAGTTTCATCTACTGTGGGTGCCG...GAGACCAAGGG... 2005

516 GlyLeuGlnGluSerSerHisArgLutrrpGlyGln 527

2006AAGGCTTAACGCCGACGAAAGTTTGGGAAGG 2035

821 GACCTTCTCACACCCCATTTTATAGATGGACAACTTTCCTCTCTGTGGTTTAT 772
323 eTrpPheSerAsnAlaPheSerTyrTyrGlyLeuValLeuLeuThrThrG 340
771 ATGGTTTCCAAATGCATTCTCTACTACGGTGTAGTTCTACTCACACAG 722
340 LuLeuPheGlnAlaGlyAspValCysGlyIleSerSerArgLysLysAla 356
721 MACTCTTCCAGGACGAGATGCTCGGGGCTCTCCAGTCGGAAGAGGCT 672
357 ValGluAlaLysCysSerLeuAlaCysGluTyrLeuSerGluGluAspTy 373
671 GTAGAGGCAAAATGACGCTCGGCTCGAGTACCTGAGTACGAGGAGTSA 622
373 rMetAspLeuLeuThrThrLeuSerGluPheProGlyValLeuVal 390
621 CATGGACTTCTCTGACGACCATCTCTCAGTTTCCAGGTGCTTGTGA 572
390 hrLeuTrpIleLeuAspArgLeuGlyArgLysLysThrMetAlaLeuCys 406
571 CTCGTGGATTAATGACCGCTTGGGGCGAAGAAGACCATGGCCCTGTGC 522
407 PheValIlePheSerPheCysSerLeuLeuLeuPheIleCysValGlyAr 423
521 TTTGTCTATCTCTCTCTGACGCTCTCTGCTGCTGTTATCTGTGTGGAAG 472
423 gAsnValLeuThrLeuLeuPheIleAlaArgAlaPheIleSerGlyG 440
471 AAATGTGCTTACTCTCTTACTCTTCTATGCAAGAGCGTTTATTTCTGGAG 422
440 LypheGlnAlaAlaTyrValTyrProGluValTyrProThrAlaThr 456
421 GCTTTTCAAGCGGATATGTTTACACACCTGAGGCTCTACCCACGGAACG 372
457 ArgAlaLeuGlyLeuGlyThrCysSerGlyMetAlaArgValClyAlaLe 473
371 CGGGCCCTCGGCTGGGACCTGCAGCGCATGGCAAGAGTGGTGTCT 322
473 uIleThrProPheIleAlaGlnValMetLeuGluSerSerValTyrLeuT 490
321 CANTACTCGGTTTCATCGCCAGGTGTATGCTGGAATCCTCTGTGTACCTGA 272
490 hrLeuAlaValTyrSerGlyCysCysLeuLeuAlaLeuAlaSerCys 506
271 CTCTGCACTTTACAGTGGCTGCTGGCTCTCTGGCTGCTGGCTGCTGCTG 222
507 PheLeuProIleLeuThrLysGlyGlyLeuGlnGluSerSerHisar 523
221 TTTTGGCCCATTTGAGACCAAGGCGGAGGACTGCAGGAGTCCAGCCACCG 172
523 gGluTrpGlyGlnGluMetValGlyArgGlyMetHisGlyAlaGlyValT 540
171 GGAGTGGGCGCAGGAGATGCTCGGCGGAGGATGCACGGTGCAGGTGTTA 122

seq_name: gb_est1:AL523336

seq_documentation_block:
LOCUS AL523336 1022 bp mRNA EST 13-FEB-2001
DEFINITION AL523336 LTI_NFL003_NBC3 Homo sapiens cDNA clone CS0DC001YN02 5
prime_mRNA sequence.
ACCESSION AL523336
VERSION AL523336.1 GI:12786829
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1022)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.

TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
location/Qualifiers
1..1022
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DC001YN02"
/clone_lib="LTI_NFL003_NBC3"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"
BASE COUNT 180 a 268 c 326 g 243 t 5 others
ORIGIN

alignment_scores:
Quality: 1256.50 Length: 258
Ratio: 5.006 Gaps: 1
Percent Similarity: 97.287 Percent Identity: 96.899

alignment_block:
US-09-911-667A-2 x AL523336 ..

Align seg 1/1 to: AL523336 from: 1 to: 1022

1 MetGluGluAspLeuPheGlnLeuArgGlnLeuProValValLysPheAr 17
231 ATGGAGGAGGACTTATTCAGCTAAGCAGCTGCCGGTTGTGAATTCG 280
17 gargThrGlyGluSerAlaArgSerGluAspThrAlaSerGlyGluH 34
281 TCGCACAGGCGAGGTGCAAGGTCAGAGGACGACACGGCTTCAGGAGG 330
34 isGluValGlnIleGluGlyValHisValGlyLeuGluAlaValGluLeu 50
331 ATGAAGTCCAGATTGAWGGGTCCACCTGGGCTAGAGGCTGTGGAGCTG 380
51 AspAspGlyAlaAlaValProLysGluPheAlaAsnProThrAspAspTh 67
381 GATGATGGGCGAGCTGTGCCAAGGAGTTTGCCAATCCACCGCATGATAC 430
67 rPheMetValGluAspAlaValGluAlaIleGlyPheGlyLysPheGlnT 84
431 TTTTCATGGTGGAGATGTCAGTGGAGGCCATTTGGAAATTTTCAGT 480
84 rPLysLeuSerValLeuThrGlyLeuAlaTrpMetAlaAspAlaMetGlu 100
481 GGAAGCTGCTCTCTCTACTGCTTGGCTTGGCTGCTGCTGCTGCTGCTG 530
101 MetMetIleLeuSerIleLeuAlaProGlnLeuHisCysGluTrpArgLe 117
531 ATGATGATCTCTCAGCATCTGGCACCCACAGTGCATTTGGAGTGGAGGT 579
117 uProSerTrpGlnValAlaLeuLeuThrSerValPheValGlyMetM 134
580 .CCAACTGGCGAGGTGGCATTCCTGACCTCGGTGGTGTGTGTAGGATGA 629
134 eLSerSerSerThrLeuTrpGlyAsnIleSerAspGlnTyrGlyArgLys 150
|||||

630 TGTCCAGCTCCAGGCTCTGGGAAATATCTCAGACCAGTACGGCAGGAAA 679
151 ThrGlyLeuIleSerValLeuThrTrpThrLeuTyrGlyIleLeuSe 167
680 ACAGGGCTGAAGATCAGCGTCTGGACTCTGTACTATGGCATCCCTAG 729
167 rAlaPheAlaProValTyrSerTrpIleLeuValLeuArgGlyLeuValG 184
730 TGCATTTCGCCCGCTGTATAGCTGGATCTTGGTCTCCGGGGCTGGTG 779
184 lypheGlyIleGlyValproGlnSerValThrLeuTyrAlaGluPhe 200
780 GCTTCGGGATCGGAGGATTCGCCAGTCGGTGCAGCTGTATGCCGAGTTC 829
201 LeuProMetLysAlaArgAlaLysCysIleLeuLeuIleGluValPheTr 217
830 CTTCCTCCATGAWGCCAGAGCTAATGATTTGCTGATGAGGTATTCG 879
217 pAlaIleGlyThrValPheGluValValLeuAlaValPheValMetPro 234
880 GGCCATCGGACAGTGTTCGAGTCGCTGGCTGTGTTCGTGATGCCCA 929
234 erLeuGlyTyrArgTrpLeuLeuIleLeuSerAlaValProLeuLeuLeu 250
930 GCCTGGGCTGGCTGTGGCTGCTCATC...TCTCAGCTGTCGCGCTCCCTC 976
251 PheAlaValLeuCysPheTrpLeu 258
977 TTTGCCGTGCTGTGTTCTGGCTG 1000

seq_name: gb_est2:B1341271

seq_documentation_block:
LOCUS B1341271 570 bp mRNA EST 30-JUL-2001
DEFINITION 368657 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION B1341271
VERSION B1341271.1 GI:15034560
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 570)
Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,F.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
and Keele,J.W.
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
Unpublished (2000)
Contact: Smith TPI
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithemail.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCCAGTCAGCAG
Plate: 108 row: D column: 4
Seq primer: ATTTAGGTGACACTATAG.
Location/Qualifiers
1..570
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 2P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/note="vector: pCMV SPORT6; site_1: XbaI; site_2: XhoI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."

FEATURES
source

seq_name: gb_est2:BG703339
seq_documentation_block:
LOCUS BG703339 638 bp mRNA EST 07-MAY-2001
DEFINITION 602685286F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4817829 5',
mRNA sequence.
ACCESSION BG703339
VERSION BG703339.1 GI:13975571
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

alignment_scores:
Quality: 986.00 Length: 190
Ratio: 5.189 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 97.895
alignment_block:
US-09-911-667A-2 x B1341271 ..
Align seg 1/1 to: B1341271 from: 1 to: 570

207 AlalysCysIleLeuLeuIleGluValPheTrpAlaIleGlyThrValph 223
1 GCTAAGTGCATTTTGGCTGATTGAGGTGTCTGGGCCATCGGGACGGTGT 50
223 eGluValValLeuAlaValPheValMetProSerLeuGlyTrpArgTrpL 240
51 CGAGGTTCATCCTGGCTGTGTGTGTATGCCAGCCTGGGCTGGCGTTGGC 100
240 euLeuIleLeuSerAlaValProLeuLeuLeuPheAlaValLeuCysPhe 256
101 TGCTCCTCTTTCAGCTGTCCCGCTCCTCTCTCTCTCTCTCTCTCTCT 150
257 TrpLeuProGluSerAlaArgTyrAspValLeuSerGlyAsnGlnGlu 273
151 TGGCTGCCAGAGAGTGCAGATATGATGTCTCTCTGGAACACAGGAAAA 200
273 sAlaIleAlaThrLeuLysArgIleAlaThrGluAsnGlyAlaProMetP 290
201 GGCCATTGCCACCTTAAAGAGGATAGCCACGAAACGGAGCCGCCATGC 250
290 roLeuGlyLysLeuIleSerArgGlnGluAspArgGlyLysMetArg 306
251 CTCTGGGGAAGCTCATCATCTCCAGACAGGAAGACCGAGGCAAAATGAG 300
307 AspLeuPheThrProHisPheArgTrpThrThrLeuLeuLeuTrpPhe 323
301 GACCTTTTCACACCCCATTTTAGATGGACACCTTTTGTGCTGTTTAT 350
323 eTrpPheSerAsnAlaPheSerTyrGlyLeuValLeuLeuThrThrG 340
351 ATGTTTTTCCAATGCGTTTCTTACTATGGACTAGTCTGCTCACCCTG 400
340 luLeuPheGlnAlaGlyAspValCysGlyIleSerSerArgLysLysAla 356
401 AGCTCTTCCAGGCTGGAGATCTCTGCAGCATCTCCAGCCGGAAGAGCC 450
357 ValGluAlaLysCysSerLeuAlaCysGluTyrLeuSerGluGluAspTy 373
451 GTAGAGGCAAAATGCAGCCTGGCCTGCAGTACCTGAGTAAGGAGGATTA 500
373 rMetAspLeuLeuTrpThrThrLeuSerGluPheProGlyValLeuVal 390
501 CATGGACCTCTGTGGACCAAGCTCTCTGAGTTTCCAGGTGCTCTTGGA 550
390 hrLeuTrpIlelleAspArg 396
551 CTCTCTGGATTATCGACCGC 570

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 638)
 NIH-MGC http://mqc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NIHRI), Shiraki
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1AM10718 row: 1 column: 22
 High quality sequence stop: 638.

FEATURES

source
 1..638
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4817829"
 /clone_lib="NIH_MGC_95"
 /tissue_type="hippocampus"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pBluescriptR (modified
 pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
 size-selected for average insert size 2.5 kb and
 normalized to ROT 5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NIH/NIHRI, National
 Institutes of Health). Note: this is a NIH_MGC Library."
 133 a 149 c 212 g 144 t

BASE COUNT

ORIGIN

alignment_scores:

Quality: 944.00 Length: 208
 Ratio: 4.697 Gaps: 5
 Percent Similarity: 96.635 Percent Identity: 95.192

alignment_block:

US-09-911-667A-2 x BG703339 ..
 Align seg 1/1 to: BG703339 from: 1 to: 638
 1 MetGluGluAspLeuPheGlnLeuArgGlnLeuProValValtysPheAr 17
 25 ATGGAGGAGGACTTATTCAGCTAAGGCAGCTGCCGGTTGTCAATTCGG 74
 17 gArgThrGlyGluSerAlaArgSerGluAspAspThrAlaSerGlyGluH 34
 75 TCGCACAGCGGAGAGTGCAGAGTCAGAGGACACACGGCTTCAGGAGAC 124
 34 IsGluValcInleGluGlyValHisValGlyLeuAlaValcInleu 50
 125 ATGAAGTCCAGATTGAAGGGTCCACGGTGGCCCTAGAGCGTGTGAGCTG 174
 51 AspAspGlyAlaAlaValProLysGluPheAlaAsnProThrAspAspTh 67
 175 GATGATGGGCGAGCTGTGCCCAAGAGAGTTGCCAATCCCACTGATGATAC 224
 67 rPheMetValGluAspAlaValGluAlaIleGlyPheGly.LysPheGln 83
 225 TTTCATGGTGAAGATGACAGTGAAGCCATTGGCTTTGGACAAATTCAG 274
 84 TrpLysLeuSerValLeuThrGlyLeuAlaTrpMetAlaAspAlaMetGl 100
 275 TGGAGAGCTGCTGTCTCAGCTGGCTTGGCTGGATGGCTGATGCATGGA 324

100 uMetMetIleLeuSerIleLeuAlaProGlnLeuHisCysGluTrpArgL 117
 325 GATGATGATCCTCAGCATCTGGCACACAGCTGCATTCGAGTGGAGGC 374
 117 euProSerTrpGlnValAlaLeuThrSerValValPheValGlyMet 133
 375 TCCCAAGCTGGCAGGTGGCATTGCTGACCTCGGTGCTTTGTAGGCATG 424
 134 MetSerSerSerThrLeuTrpGlyAsnIleSerAspGlnTrpGlyArg.L 150
 425 ATGTCCAGCTCCACGCTCTGGGGAATATCTCAGACCAGTACGCGAGGA 474
 150 ysThrGlyLeuLysIleSerValLeuTrpThrLeuTyr.TyrGlyIleLe 166
 475 AAACAGGCTGACGATCAGCTGCTGGACTCTGTACCTATATGGCATCCT 524
 166 uSerAlaPheAlaProValTyrSerTrpIleLeuValLeuArgGlyLeuV 183
 525 TAGTGCATATGCGCGGTG.TATAGCTGGATCCTGGTCCCGGGCGCTGG 573
 183 aiGlyPheGlyIle.GlyGlyValProGln.SerValThrLeuTyrAlaG 199
 574 TGGGCTTCGGGATACGAGGAGGTTCGCCAAGTCGGTGACGCTGATGCCG 623
 199 lupheLeuProMet 203
 624 AGTACTTCCCATG 637
 seq_name: gb_est2:BI394002
 seq_documentation_block:
 LOCUS BI394002 662 bp mRNA EST 06-AUG-2001
 DEFINITION pgpln.pk013.c4 Normalized Chicken Pituitary/Hypothalamus/Pineal
 Library Gallus gallus cDNA clone pgpln.pk013.c4 5' similar to
 gb|AAC78627.1 (AF060173) SV2 related protein [Rattus norvegicus],
 mRNA sequence.
 ACCESSION BI394002
 VERSION BI394002.1 GI:15087284
 KEYWORDS EST.
 SOURCE chicken.
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 1 (bases 1 to 662)
 Porter, T.E. and Cogburn, L.A.
 ESTs from Normalized Chicken Pituitary/Hypothalamus/Pineal cDNA
 Library USDA/IFAFS Animal Genome Project
 Unpublished (2001)
 Contact: Larry A. Cogburn
 University of Delaware
 Townsend Hall, Newark, DE 19717, USA
 Tel: 302-831-1335
 Fax: 302-831-2822
 Email: cogburnudel.edu, www.chickest.udel.edu.
 FEATURES
 source
 1..662
 /organism="Gallus gallus"
 /strain="Commercial broiler chicken"
 /db_xref="taxon:9031"
 /clone="pgpln.pk013.c4"
 /clone_lib="Normalized Chicken
 Pituitary/Hypothalamus/Pineal Library"
 /sex="Male and Female"
 /tissue_type="Pituitary Gland/Hypothalamus/Pineal Gland"
 /dev_stage="Embryonic (d12,d14,d19); post-hatch (w1,w3,w5,
 w7,w9)"
 /lab_host="E. Coli BMDH10B"
 /note="Vector: pCMVSPORT6; Library made from equivalent
 pools of total RNA isolated from each tissue at different
 ages. Single pass sequencing from 5'-end"
 160 a 145 c 169 g 177 t 11 others

BASE COUNT

ORIGIN

```
alignment_scores:
  Quality: 939.00      Length: 199
  Ratio: 4.916        Gaps: 0
Percent Similarity: 95.980 Percent Identity: 90.955

alignment_block:
US-09-911-667A-2 x BL394002 ..

Align seg 1/1 to: BL394002 from: 1 to: 662

350 IieSerSerArgLysLysAlaValGluAlaLysCysSerLeuAlaCysG1 366
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
6 ATTTCCAGTAGAAGAAAGAAAGCTTAAAGCAAGAGTGACGCTGACCTGTCA 55

366 uTyrLeuSerGluGluAspTyrMetAspLeuLeuTrpThrLeuSerG 383
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
56 GTATCTGACAGAGGAGACTACACTGATCTGCTTGGACAAACCTGTCCAG 105

383 IuPheProGlyValLeuValThrLeuTrpIleIleAspArgLeuGlyArg 399
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
106 AATTCCTCCAGGTGTGTAGTAACACTGTGGATTATTGATCGGATAGCCCG 155

400 LysLysThrMetAlaLeuCysPheValIlePheSerPheCysSerLeuLe 416
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
156 AAGAAACCATGGCCCTGCTCTTCTTGTCTTCATTTTGCAGCCTGCT 205

416 uLeuPheIleCysValGlyArgAsnValLeuThrLeuLeuPheIleA 433
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
206 GCTGTTTCTCTCTGTGTGAAGAAATGTTTACTGTGCTGCTTTCATTTG 255

433 IaArgAlaPheIleSerGlyGlyPheGlnAlaAlaTyrValTyrThrPro 449
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
256 CAAGAGCTTTTATTTCAGAGGATTTACAGCTGCTTATGTTTACACTCCG 305

450 GluValTyrProThrAlaThrArgAlaLeuGlyLeuGlyThrCysSerG1 466
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
306 GAGGTTTACCAACAGCCAGCGCTGCTTGGGCTGGGAACATGCAGTGG 355

466 yMetAlaArgValGlyAlaLeuIleThrProPheIleAlaGlnValMetL 483
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
356 AATGGCCAGAGTGGAGCCCTCATACACCATTCATTGACACAGGTATGT 405

483 euGluSerSerValTyrLeuThrLeuAlaValTyrSerGlyCysCysLeu 499
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
406 TGAATCTTCAGTCTATTAAACACTGGCAGTTACAGTGATGTGCCCAG 455

500 LeuAlaLeuAlaLeuSerCysPheLeuProIleGluThrLysGlyGlyG1 516
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
456 CTGGCTGCTGTGGCTTCTCTTCTTCTTCCCATTTGAAACAAAGGTCGTG 505

516 yLeuGlnGluSerSerHisArgGluTrpGlyGlnGluMetValGlyArgG 533
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
506 CTTCAGGAGTCCAGCCACAGAGAATGGGACAGAGAGTGGTTGGAGAG 555

533 lyMetHisGlyAlaGlyValThrArgSerAsnSerGlySerGlnGlu 548
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
556 GATCTCATTCCTCCAGAGTACCAGCTTAACTCTGGGTGCACAGGAG 602

seq_name: gb_est1:AL529154

seq_documentation_block:
LOCUS AL529154 778 bp mRNA EST 13-FEB-2001
DEFINITION AL529154 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DD004YD17 5
prime, mRNA sequence.
ACCESSION AL529154
VERSION AL529154.1 GI:12792647
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE
1 (bases 1 to 778)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..778
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DD004YD17"
/clone_lib="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 150 a 193 c 254 g 166 t 15 others
ORIGIN

alignment_scores:
  Quality: 934.00      Length: 199
  Ratio: 4.968        Gaps: 1
Percent Similarity: 94.472 Percent Identity: 93.467

alignment_block:
US-09-911-667A-2 x AL529154 ..

Align seg 1/1 to: AL529154 from: 1 to: 778

1 MetGluGluAspLeuPheGlnLeuArgGlnLeuPro.ValValLysPhea 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
182 ATGGAGGAGGACTTATTCAGCTAAGCAGCCGCCGGTTGTGAATCC 231

17 rArgThrGlyGluSerAlaArgSerGluAspThrAlaSerGlyGlu 33
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
232 GTCCACAGGCGAGAGTGCAGGTCAGAGGACGACACGGCTTCAGGAG 281

34 HisGluValGlnIleGluGlyValHisValGlyLeuGluAlaValGluLe 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
282 CATGAAGTCCAGATTGAAGGGTCCACGTGGGCTTAGAGGCTGTGGAGCT 331

50 uAspAspGlyAlaAlaValProLysGluPheAlaAsnProThrAspSpt 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
332 GCATGATGGGCGAGCTGTGCCCAAGGAGTTTGCCAAATCCACCGATGATA 381

67 hrPheMetValGluAspAlaValGluAlaIleGlyPheGlyLysPheGln 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
382 CTTTCATGTTGGAAGATGCAGTGGGAAGCCATTGGCTTTGGAAAATTTTCAG 431

84 TrpLysLeuSerValLeuThrGlyLeuAlaTrpMetAlaAspAlaMetG1 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
432 TGGAAAGTCTCTGTCTCACTGGCTTGGCTTGGATGGCTGATGCCATGGA 481

100 uMetMetIleLeuSerIleLeuAlaProGlnLeuHisCysGluTrpArgL 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
482 GATGATGATCCTCAGCATCTGTCACACAGCTGCTTGCAGAGTGAGGC 531

117 euProSerTrpGlnValAlaLeuLeuThrSerValValPheValGlyMet 133
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
532 TCCCAAGTGGCAGGTGATGCTGCTGAMCTCGTGGTGTCTTTTAGGCATG 581
```

134 MetSerSerSerThrLeuTrpGlyAsnIleSerAspGlnTyrGlyArgLy 150
 |||||
 582 AMGTCAGCTCCACGCTCTGGGAAATATYTCAGACCAGTACGGCAGGAA 631
 |||||
 150 sThrGlyLeuLysIleSerValLeuTrpThrLeuTyrGlyIleLeuS 167
 |||||
 632 AACARGGCTGAAGATCAGCGTCTGTGACTCTKTCWTATGCACTCCCTTA 681
 |||||
 167 erAlaPheAlaProValTyrSerTrpIleLeuValLeuArgGlyLeuVal 183
 |||||
 682 GTGCATTTGGCGCCGCTKTWAGCTGGATCCTGGTCTCCGGGGCCTGGTG 731
 |||||
 184 GlyPheGlyIleGlyGlyValProGlnSerValThrLeuTyrAla 198
 |||||
 732 GGCTTCGGGATCGAGAGAGTTCGCCAGTCGGTGACGCTKTTTGCC 776
 |||||

seq_name: gb_est1:AL563357

seq_documentation_block: 642 bp mRNA EST 16-FEB-2001
 LOCUS AL563357 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DD004YD17 3
 DEFINITION prime, mRNA sequence.

ACCESSION AL563357

VERSION AL563357.1 GI:12912671

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Li, W.B., Gruber, C., Jesse, J., and Polayes, D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

source

1. 642

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CS0DD004YD17"

/clone_lib="LTI_NFL001_NBC4"

/sex="male"

/tissue_type="neuroblastoma cells"

/lab_host="DH10B"

/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA

was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-stranded cDNA was digested with Not I and

cloned into the Not I and Eco RV sites of the pCMVSPORT 6

vector. Library was normalized. Library was constructed

by life technologies. Contact : Feng Liang Life

Technologies, a division of Invitrogen 9800 Medical Center

Drive Rockville, Maryland 20850, USA Fax : (1) 301 610

8371 Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com"

BASE COUNT 149 a 192 c 170 g 117 t 14 others

ORIGIN

alignment_scores:

Quality: 898.00 Length: 202

Ratio: 4.854 Gaps: 1

Percent Similarity: 91.584 Percent Identity: 90.099

alignment_block:

US-09-911-667A-2 x AL563357/rev ..

Align seg 1/1 to reverse of: AL563357 from: 1 to: 642

348 CysGlyIleSerSerArgLysLysAlaValGluAlaLysCysSerLeuAl 364

|||||

637 TCGCGCTTCTCCAGTCGGAARAAGCTGAAGAGCGCAAAATGACGCTGCG 588

364 aCysGlyTyrLeuSerGluGluAspTyrMetAspLeuLeuTrpThrThrL 381
 |||||
 587 CTGCGAGTACCTGAGTGAGGAGGATWACATGGACTTGTGTGACCAACCC 538
 |||||
 381 euSerGluPheProGlyValLeuValThrLeuTrpIlelleAspArgLeu 397
 |||||
 537 TCTCTGAGTTTCCAGGTGCTTTGTGACTCTGTGGATTATTGACCGCTG 488
 |||||
 398 GlyArgLysLysThrMetAlaLeuCysPheValIlePheSerPheCysse 414
 |||||
 487 GGGCGCAAGAACACCATGGCCCTGGGCTTTGTGCATCTTCTCTTCNGCAG 438
 |||||
 414 rLeuLeuLeuPheIleCysValGlyArg, AsnValLeuThrLeuLeuLeu 430
 |||||
 437 CCNCCNGCTGTTTNNCTGTGGAAGAAATGNGCTCACTCTGTAACCTC 388
 |||||
 431 PheIleAlaArgAlaPheIleSerGlyGlyPheGlnAlaAlaTyrValTy 447
 |||||
 387 TTTCATGGCAAGAGCGTTTATTCTTGAGGCTTTTCAAGCGGCATATGTTNA 338
 |||||
 447 rThrProGluValTyrProThrAlaThrArgAlaLeuGlyLeuGlyThrC 464
 |||||
 337 CACACCTGAGGTCTACCCACGCAACGGCGGCTCGGCCCTGGCGTGGGCACCT 288
 |||||
 464 ysSerGlyMetAlaArgValGlyAlaLeuIleThrProPheIleAlaGln 480
 |||||
 287 GCAGCGGCATGGCAAGAGNGGGGCTCTCANTACTCCGTTTCATCGCCAG 238
 |||||
 481 ValMetLeuGluSerSerValTyrLeuThrLeuAlaValTyrSerGlyCy 497
 |||||
 237 GTGATGCTGGAATCTCTGTGTACCTGACTCTGGCAGTTAACAGTGGCGNG 188
 |||||
 497 sCysLeuLeuAlaAlaLeuAlaSerCysPheLeuProIleGluThrLysG 514
 |||||
 187 CTGCTCTCTGGCTGCGCTGCGCTCTCTGCTTTTGGCCATTGAGACCAAG 138
 |||||
 514 lylGlyGlyLeuGlnGluSerSerHisArgGluTrpGlyGlnGluMetVal 530
 |||||
 137 GCCGAGNACTGCGAGGAGTCCAGCCACGGAGTGGGGCCAGGAGATGGCTC 88
 |||||
 531 GlyArgGlyMetHisGlyAlaGlyValThrArgSerAsnSerGlySerGI 547
 |||||
 87 GGGCGAGGAATGACGGTGCAGGTGTTTACAGGTGCGGACTCTGGCTCTCA 38
 |||||
 547 nGlu 548
 |||||
 37 GGAA 34

seq_name: gb_est2:BF775076

seq_documentation_block:

LOCUS BF775076 468 bp mRNA EST 25-APR-2001

DEFINITION 285128 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.

ACCESSION BF775076

VERSION BF775076.1 GI:12122976

KEYWORDS EST.

SOURCE cow.

ORGANISM Bos Taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 468)

AUTHORS

Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,

Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett

, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G.,

Pertea, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J., and

Keele, J.W.

TITLE

Sequence evaluation of four pooled-tissue normalized bovine cDNA

libraries and construction of a gene index for cattle

JOURNAL

Genome Res. 11 (4), 626-630 (2001)

MEDLINE

21180013

COMMENT

Co tact: Smith TPL

USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.

PCR Primers
FORWARD: AGGAACACGTATGACCAT
BACKWARD: GTTTTCCCGATCAGCAGC
Plate: 86 row: E column: 5
Seq primer: ATTTAGCTGACACTATAG.

FEATURES

source
1. .468
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 3BOV"
/tissue_type="pooled"
/lab_host="DH108"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."
BASE COUNT 87 a 126 c 139 g 116 t
ORIGIN
alignment_scores:
Quality: 809.00 Length: 156
Ratio: 5.186 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-911-667A-2 x BF775076 ..
Align seg 1/1 to: BF775076 from: 1 to: 468

163 TyrGlyIleLeuSerAlaPheAlaProValTyrSerTrpIleLeuValLe 179
1 TACGGCATCCCTCAGTGCCTTTGGCCCGGTACAGCTGGATCCTGGTGT 50
179 uArgGlyLeuValGlyPheGlyIleGlyValProGlnSerValThrL 196
51 CGGGGCGCTGTGGCTTCGGAGTCGGAGGGTCCCCAGTCGGTCACAC 100
196 euTyrAlaGluPheLeuProMetLysAlaArgAlaLysCysIleLeuLeu 212
101 TGTATGCTGAGTTCCTTCCCATGAAAGCCAGAGCTAAATGATTTTGTG 150
213 IleGluValPheTrpAlaIleGlyThrValPheGluValValLeuAlaVa 229
151 ATTGAGGTGTTCTGGCCCATCGGACAGTGTTCGAGGTGCTCTGGCTGT 200
229 IpheValMetProSerLeuGlyTrpArgTrpLeuLeuIleLeuSerAlaV 246
201 GTTTGTGATGCCACCGCTGGCTGGCTGGCTGCTCATCTCTCAGCTG 250
246 alProLeuLeuPheAlaValLeuCysPheTrpLeuProGluSerAla 262
251 TCCCACTCCTCTCTTTCGCTGCTGTGTGTTTGGCTGCCGAGAGTGGC 300
263 ArgTyrAspValLeuSerGlyAsnGlnGluLysAlaIleAlaThrLeuLy 279
301 AGGTATGATGTGCTGTCCGGAACACGAGAGAGCCATCGCCACCTTAA 350
279 sArgIleAlaThrGluAsnGlyAlaProMetProLeuGlyLysLeuLeI 296
351 GAGAAATAGCCAGGAAACGAGAGCTCCATGCCCTCTGGGAAACTCATCA 400
296 leSerArgGlnGluAspArgGlyLysMetArgAspLeuPheThrProHis 312
401 TTTCCAGACGAAAGACCGGAGGCAAAATGAGGGACCTTTTTCACACCC 450

313 PheArgTrpTrpThrLeu 318
|||||
451 TTTAGATGCACAACCTTG 468
seq_name: gb_est1:AW173250
seq_documentation_block:
LOCUS AW173250 487 bp mRNA EST 16-NOV-1999
DEFINITION x185b11.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2663997 3' similar to TR:Q92217 Q92217 SV2 RELATED PROTEIN.
.: mRNA sequence.

ACCESSION AW173250
VERSION AW173250.1 GI:6439198
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 487)
REFERENCE
AUTHORS
TITLE
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
COMMENT
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cygabs-remail.nih.gov

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Trace considered overall poor quality
Seq primer: -40UP from Gibco
High quality sequence stop: 1.

Location/Qualifiers

1. .487
source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2663997"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH108"
/note="Organ: pooled; Vector: p7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NbHL19W, testis NHT, and B-cell
NCI CGAP GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT 126 a 143 c 133 g 84 t 1 others
ORIGIN

alignment_scores:

Quality: 760.00 Length: 162
Ratio: 4.872 Gaps: 0
Percent Similarity: 96.296 Percent Identity: 90.741

alignment_block:

US-09-911-667A-2 x AW173250/rev ..

Align seg 1/1 to reverse of: AW173250 from: 1 to: 487

386 GlyValLeuValThrTrpIleIleAspArgLeuGlyArgLysLysTh 402
|||||
487 GGTGTCATGTGACTCTGTGGATTATTCACCGCATGGGGCGCAAGAC 438
402 rMetAlaLeuCysPheValIlePheSerPheCysSerLeuLeuPheI 419
|||||
437 CATGGCCNTGTGCTGTGCATCTTCTCTTGTGCAGCTCGTGTGTTTA 388
419 leCysValGlyArgAsnValLeuThrLeuLeuPheIleAlaArgAla 435

```

|||||.....:|||||.....
387 TCTGTGTTGGAGAGATGCTGCTACTCTGTACTCTTCATTGCAAGAGCG 338
436 PheIleSerGlyGlyPheGlnAlaIaIaValTyrThrProGluValTy 452
|||||.....:|||||.....
337 TTTATTATGAGGCTTTCAACGGGCATATGTTTACACACGTGAGGTCTA 288
452 rProThrAlaThrArqAlaLeuGlyLeuGlyThrCysSerGlyMetAlaA 469
|||||.....:|||||.....
287 CCCACGGCAACGGGGCCCTCGGCTGGCAGCTGCAGCGGCGCA 238
469 rGValGlyAlaIaLeuIleThrProPheIleAlaGlnValMetLeuGluSer 485
|||||.....:|||||.....
237 GAGTGGGTGCTCTCATCTCCGTTTCATCCCGCAGGTGCTGGAATCC 188
486 SerValTyrLeuThrLeuAlaValTyrSerGlyCysCysLeuLeuAlaI 502
|||||.....:|||||.....
187 TCTGTGTACCTGACTCTGGCAGTTTACAGTGGCTGCTGCTGCTGCTGC 138
502 aLeuAlaSerCysPheLeuProIleGluThrLysGlyGlyLeuGlnG 519
|||||.....:|||||.....
137 CCTGGCTCTCTGCTTTTGGCCATTGAGACCAAGGCCGAGGACTGCAGG 88
519 luSerSerHisArqGluTrpGlyGlnMetValGlyArgGlyMetHis 535
|||||.....:|||||.....
87 AGTCCACCCACGGGAGTGGGGCCAGAGATGTCGGCCCAAGGAATGCAC 38
536 GlyAlaGlyValThrArgSerAsnSerGlySerGln 547
|||||.....:|||||.....
37 GGTTCAGCTGTTCCAGGTGCAACTCTGGCTCTCAG 2

```

seq_name: gb_est1:AW663941

seq_documentation_block: 578 bp mRNA EST 06-APR-2000
LOCUS AW663941
DEFINITION hi73h10.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2977987 3' similar to TR:Q92217 Q92217 SV2 RELATED PROTEIN.
: mRNA sequence.

ACCESSION AW663941
VERSION AW663941.1 GI:7456480

KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 578)

NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cyapbs-r@mail.nih.gov

This clone is available royalty-free through LINL: contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

Seq primer: -40UP from Gibco

High quality sequence stop: 464.

FEATURES

Location/Qualifiers

source

1..578

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2977987"

/clone_lib="Soares_NFL_T_GBC_S1"

/lab_host="DH10B"

/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with

a modified polylinker; Site_1: Not 1; Site_2: Eco RI;

Equal amounts of plasmid DNA from three normalized

libraries (fetal lung NbHL19W, testis NHT, and B-cell

NCI_CGAP_GCB1) were mixed, and ss circles were made in

vitro. Following HAP purification, this DNA was used as

tracer in a subtractive hybridization reaction. The driver

was PCR-amplified cDNAs from pools of 5,000 clones made

from the same 3 libraries. The pools consisted of

I.M.A.G.E. clones 297480-302087, 682632-687239,

726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo.

BASE COUNT 137 a 170 c 159 g 111 t 1 others
ORIGIN

alignment_scores:

Quality: 743.00 Length: 151
Ratio: 5.020 Gaps: 1
Percent Similarity: 98.013 Percent Identity: 98.013

alignment_block:

US-09-911-667A-2 x AW663941/rev ..

Align seg 1/1 to reverse of: AW663941 from: 1 to: 578

```

399 ArgLysLysThrMet..AlaLeuCysPheValIlePheSerPheCysSerL 415
|||||.....:|||||.....
576 CGCAAAAGACCATGGGCCCTGTCTTGTTCATCTCTCTCTCTGACGCC 527
415 euleLeuPheIleCysValGlyArgAsnValLeuThrLeuLeuPhe 431
|||||.....:|||||.....
526 TCTGTGCTGTTATCTGTGTGGAGANATGTCTCACTCTGTCTTACTCTTC 477
432 IleAlaArgAlaPheIleSerGlyGlyPheGlnAlaAlaIaIaValTyrTh 448
|||||.....:|||||.....
476 ATTGCAAGAGCGTTTATTCTGGAGGCTTTCAAGCGGCATATGTTTACAC 427
448 rProGluValTyrProThrAlaThrArgAlaLeuGlyLeuGlyThrCysS 465
|||||.....:|||||.....
426 ACCTGAGGTCTACCCACGCAACGCGGCCCTCGGCCCTGGGCACCTGCA 377
465 erGlyMetAlaArgValGlyAlaLeuIleThrProPheIleAlaGlnVal 481
|||||.....:|||||.....
376 GCGGCATGCAAGAGTGGGTGCTCTCATCTCCGTTCTATCGCCCAAGTG 327
482 MetLeuGluSerSerValTyrLeuThrLeuAlaValTyrSerGlyCysCy 498
|||||.....:|||||.....
326 ATGCTGGAATCTCTGTGTACCTGACTCTGGCAGTTTACAGTGGCTGCTG 277
498 sL..uLeuAlaLeuAlaSerCysPheLeuProIleGluThrLysGlyC 515
|||||.....:|||||.....
276 CTTCTCTGGCTGGCCCTGGCCCTCTCTGCTTTTGGCCCATGAGACCAAGGCC 227
515 lGlyLeuGlnGluSerSerHisArqGluTrpGlyGlnGluMetValGly 531
|||||.....:|||||.....
226 GAGGACTGCAGGAGTCCACCCAGCGGAGTGGGGCCAGGAGATGTCGGC 177
532 ArgGlyMetHisGlyAlaGlyValThrArgSerAsnSerGlySerGlnG 548
|||||.....:|||||.....
176 CGAGGAATGCAGGTGCAGGTGTTACCCAGTGGAACTCTGGCTCTCAGGA 127
548 u 548
126 a 126

```

seq_name: gb_est1:AU220193

seq_documentation_block:

LOCUS AU220193 742 bp mRNA EST 17-JUL-2001
DEFINITION AU220193 unpublished oligo-capped cDNA library, stage LI
Caenorhabditis elegans cDNA clone yk86g01 3', mRNA sequence.

ACCESSION AU220193

VERSION AU220193.1 GI:14858350

KEYWORDS EST.

SOURCE Caenorhabditis elegans.

ORGANISM Caenorhabditis elegans.

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea

; Rhabditidae; Peloderinae; Caenorhabditis.

1 (bases 1 to 742)

REFERENCE Kohara, Y., Shin-i, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.

and Sugano, S.

A complementary view of the C.elegans genome

TITLE

JOURNAL Unpublished (2001)
 COMMENT Contact: Yuji Kohara
 Genome Biology Lab.
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan
 Tel: 81-559-81-6854
 Fax: 81-559-81-6855
 Email: ykohara@lab.nig.ac.jp.
 Location/Qualifiers
 source 1. 742
 /organism="Caenorhabditis elegans"
 /strain="N2"
 /db_xref="taxon:6239"
 /clone_lib="yk886g01"
 /clone_lib="unpublished oligo-capped cDNA
 L1"
 /sex="Hermaphrodite"
 /tissue_type="whole animal"
 /dev_stage="L1"
 BASE COUNT 247 a 154 c 140 g 201 t
 ORIGIN
 alignment_scores:
 Quality: 619.00 Length: 237
 Ratio: 3.241 Gaps: 3
 Percent Similarity: 80.591 Percent Identity: 51.899
 alignment_block:
 US-09-911-667A-2 x AU220193/rev ..
 Align seg 1/1 to reverse of: AU220193 from: 1 to: 742
 289 MetProLeuGlyLysLeuIleSeraArgln.....GluAspAr 302
 :::||| :|||::|||::||| :|| :|||::|||::|||
 739 TTGCCAACTGGAACTAGTCTTCAACAAAAGCTGGATCAGAAAGTAG 690
 302 gGlyLysMetArgAspLeuPheThrProHisPheArgTrpThrLeuL 319
 ||||| ::: :|||::|||::||| :||| :|||::|||::|||
 689 AGGAGATATTCTAATTGGTTTATCTCTGACTTGTAGAAAAACAAC 640
 319 euLeuTrpPheIleTrpPheSerAsnAlaPheSerTyrTyrGlyLeuVal 335
 ||||| :||| :|||::|||::||| :|||::|||::|||
 639 TATTATGGTGCATTGGGCTATTACCCTATTCTGACTACGGTATGGTT 590
 336 LeuLeuThrThrGluLeuLeuPheGlnAlaGlyAspValCys.....GlyI 350
 :|||::|||::||| :|||::|||::||| :||| :|||::|||::|||
 589 CTCTTCACACAGGTTTGTCTTCAATCACACGATGAATGTCATGGTGA 540
 350 eSerSerArgLysLysAlaValGluAlaLysCysSerLeuAlaCysGlu 367
 : :|||::|||::||| :||| :|||::|||::|||
 539 GTTCTCTAATGGAACTCAATGGAA.....GTATGCCAAC 505
 367 yrLeuSerGluGluAspTyrMetAspLeuLeuLeuTrpThrLeuSerGlu 383
 |||::|||::||| :|||::|||::||| :|||::|||::|||
 504 CATTCACAAGATCTGATTATTTCATTATTATCACTACACTTGC CGAA 455
 384 PheProGlyValLeuValThrLeuTrpIleIleAspArgLeuGlyArg 400
 :|||::|||::||| :|||::|||::||| :||| :|||::|||::|||
 454 TTCCCTCGATTGATTAATCACTGATTGAATTAATGAATGTTTGGAA 405
 400 sLysThrMetAlaLeuCysPheValIlePheSerPheCysSerLeuLeu 417
 :|||::|||::||| :|||::|||::||| :|||::|||::|||
 404 GAACACAATGGCATTTGGAATATACGATTTTTCGAATTTTCACCTTC 355
 417 euPheIleCysValGlyArgAsnValLeuThrLeuLeuLeuPheIleAla 433
 :|||::|||::||| :||| :|||::|||::||| :|||::|||::|||
 354 TATACTTTTCTGGATCGTTTTCACGTGCTGCTCATTTTGTGGCA 305
 434 ArgAlaPheIleSerGlyGlyPheGlnAlaAlaTyrValTyrThrPro 450
 :|||::|||::||| :|||::|||::||| :||| :|||::|||::|||
 304 CGTGCCCTCATTTCTGGTGCCTTCCAACTGTCATACGCTACACACCG 255


```
Align seg 1/1 to: AA349847 from: 1 to: 371
329 pheSerTyrTyrGlyLeuValLeuLeuThrThrGluLeuPheGlnAlaG1 345
1 TTTCTTTACTAGGGTTAGTTTCTACTACACAGAACTCTTCCAGGCAGG 50
345 yaspValCysGlyIleSerArgLysLysAlaValGluAlaLysCys 362
51 AGATGCTTGGCGCATCTCCAGTCGGAAGAAGGCTCTAGAGGCAAAATGCA 100
362 erLeuAlaCysGluTyrLeuSerGluGluAspTyrMetAspLeuLeuTrrp 378
101 GCCTGGCCTGGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTAC 150
378 ThrThrLeuSerGluPheProGlyValLeuValThrLeuTrrpIleLeAs 395
151 ACCACCCCTCTGAGTTTCCAGGTCCTTGTGACTCTNTGATTTATTTGA 200
395 pArLeuGlyArgLysLysThrMetAlaLeuCysPheValIlePheSerP 412
201 CCGCTGGGGCGCAGAGACCATGGCCTGTGCTTTGTATCTTCTCTCT 250
412 heCysSerLeuLeuLeuPheIleCysValGlyArgAsnValLeuThrLeu 428
251 TCTGCAGCCTCTCTGCTGTTTATCTGTGTTGGAAGAAATGCTCACTCTG 300
429 LeuLeuPheIleAlaArgAlaPheIleSerGlyGlyPheGlnAlaIaTy 445
301 TTTACTTTTCAATGCAAGAGCGTTTATTTCTGGAGGCTTTCAAGCGGCA 350
445 rValTyrThrProGluVal 451
351 TGTTTACAAACTGAGGTC 369
seq_name: 9b_gss:CNS05LV9
seq_documentation_block:
LOCUS      CNS05LV9      1047 bp      DNA      26-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T3 end of clone
            053F06 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION  AL343278
VERSION    AL343278.1 GI:8237036
KEYWORDS   GSS; genome survey sequence.
SOURCE     Tetraodon nigroviridis.
            Tetraodon nigroviridis.
ORGANISM   Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE  1 (bases 1 to 1047)
            Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
            Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
            Unpublished
REFERENCE  2 (bases 1 to 1047)
            Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
            Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
            Unpublished
REFERENCE  3 (bases 1 to 1047)
            Genoscope.
            Direct Submission
            Submitted (12-APR-2000) to the EMBL/GenBank/DBSJ databases
            This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
            Location/Qualifiers
FEATURES
```

```
source      1..1047
            /organism="Tetraodon nigroviridis"
            /db_xref="taxon:99883"
            /clone="053F06"
            /clone_lib="A"
            /note="Genoscope sequence ID : C0AA053DC03A1-end : T3"
BASE COUNT  255 a   278 c   269 g   229 t   16 others
ORIGIN
alignment_scores:
    Quality: 587.00      Length: 275
    Ratio: 3.739        Gaps: 6
    Percent Similarity: 57.091    Percent Identity: 49.091
alignment_block:
US-09-911-667A-2 x CNS05LV9/rev
Align seg 1/1 to reverse of: CNS05LV9 from: 1 to: 1047
358 GluAlaLysCysSerLeuAlaCysGluTyrLeuSerGluGluAspTyrMe 374
881 GAGCTCCGTTGCAGCTGGAGTGAATATCTGAACTCAGACGACTACAA 832
374 tAspLeuLeuTrrpThrLeuSerGluPhePro..... 385
831 GGACCTGCTGGACCACCTTTGCTGAAATTTCCAGGCAGCTACTGGTTCC 782
385 ..... 385
781 ATGTGTCAGACAAACCTTAATTCATCCCTGATGAATTTACTGATGGCA 732
386 .....GlyValLeuValThrLeuTrrpIleAspAr 396
731 TTGTTGTTGTTTACGCGAGACTGCTGTCGACACTGGGCAATCGATCG 682
396 gLeuGlyArgLysLysThrMetAlaLeuCysPheValIlePheSerPheC 413
691 GCTGGGAGGACGAGACCATGGCTTGTGTTTCTTCATCTTCTCCATGT 632
413 ySerLeuLeuLeuPheIleCysValGlyArg.AsnVal..... 425
631 GCATTATTTCCACTCTACGGCTGTGTCGCCGGTAAGATTTCATAGCAGCTG 582
425 ..... 425
581 GTCAGATTCCAACTGGAGAGCCATGCAGNATACTGCCTTATTTCGCC 532
426 .....LeuThrLeuLeuLeuPheIleAlaArgAlaPheIleSer 438
531 GCAGGGCTCCTACCTGCTGTTGATATTCATCGCCAGAGCTTTCATTGCA 482
439 GlyGlyPheGlnAlaAlaTyrValTyrThrProGlu..Val..... 451
481 GGAGGATTTACGGCCCTNNGTTTACACTCCAGAGGTAAAGAGCCAGT 432
451 ..... 451
431 TAGAATGCAGATGACAGTTCAAGCCCATTCGCTGTTGTGCAATCTCTGCA 382
452 ....TyrProThrAlaThrArgAlaLeuGlyLeuGlyThrCysSerGlyM 467
381 GGTGTATCCACAGCGCAGGAGGCTTTGGGTCTGGGAACAAGCAGCGGAA 332
467 etAlaArgValGlyAlaLeuIleThrProPheIleAlaGln..... 480
331 TGGCAAGAGTCGGCGCCTGATCACACCATTTGTGGCAGAGGTGAGGACG 282
480 ..... 480
281 TCCTTTGCTTATTGAGCTCAAAACCTCAAAAGATGTGGTGTACACCTGC 232
481 .....ValMetLeuGluSerSerValTyrLeuThrLeuA 492
```

```

231 CGGCTTGTCTTAAAGGTGATGCTGGAGTGTGCTGCTGCTGCTGT 182
|||||
492 laValTyrSerGlyCysLeuLeuAlaLeuAlaSerCysPheLeu 508
|||||
181 TTGTGTACTGTGCTGCTGCTCTTCTGCTGCTGCTGCTGCTGCTG 132
|||||
509 ProfileGluThrLysGlyGlyLeuGluSerSerHisArgGluTr 525
|||||
131 CCCATCGAGACAGGGTGGCGGCTGCGAGAGTCCAGCCAGCAGAGT 82
|||||
525 pGlyGlnGluMetValGlyArg 532
|||||
81 GGGCCAGCAAAATGTTGGGCCGC 60
|||||

seq_name: gb_gss:CNS04275

seq_documentation_block:
LOCUS      CNS04275      868 bp      DNA      26-JUL-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
            040B23 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION AL313898
VERSION   AL313898.1 GI:9546782
KEYWORDS  GSS; genome survey sequence.
SOURCE    Tetraodon nigroviridis.
ORGANISM  Tetraodon nigroviridis.
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
           Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
           Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 868)
AUTHORS   Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
           Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
           Saurin,W. and Weissenbach,J.
           Estimate of human gene number provided by genome-wide analysis
           using Tetraodon nigroviridis DNA sequence
           Nat. Genet. 25 (2), 235-238 (2000)
20296633
REFERENCE 2 (bases 1 to 868)
AUTHORS   Crolius,H.R., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C.,
           Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W.,
           Bernot,A. and Weissenbach,J.
           Characterization and repeat analysis of the compact genome of the
           freshwater pufferfish tetraodon nigroviridis
           Genome Res. 10 (7), 939-949 (2000)
20359837
REFERENCE 3 (bases 1 to 868)
AUTHORS   Genoscope.
           Direct Submission
           Submitted (12-APR-2000) to the EMBL/GenBank/DBSJ databases
           This sequence is a single read and was generated as part of a large
           scale clone-end sequencing project of the Tetraodon nigroviridis
           genome. For more information, please take a look at
           http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source            1..868
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="040B23"
                     /clone_lib="A"
                     /note="Genoscope sequence ID : C0AA040CA12C1-end : T7"
BASE COUNT          174 a 222 c 238 g 217 t 17 others
ORIGIN

alignment_scores:
Quality: 538.50      Length: 273
Ratio: 3.614        Gaps: 5
Percent Similarity: 54.579 Percent Identity: 47.619

alignment_block:
US-09-911-667A-2 x CNS04275

```

```

Align seg 1/1 to: CNS04275 from: 1 to: 868

376 LeuLeuTrpThrThrLeuSerGluPhePro..... 385
|||||
16 CTGCTGTGGACCACTTTGTCTGAATTTCCAGGCACCTACTGTTCCATGT 65
|||||
385 ..... 385
|||||
66 GTCAGACAAACCTTAATTCATCCGCGATGAATTTACTGTGTCATTGT 115
|||||
386 ..... GlyValLeuValThrLeuTrpLleIleAspArgLeu 397
|||||
116 TGTGTTTGTGACGACGACTGCTGTGACACTGTGGCAATCGATCGGNTG 165
|||||
398 GlyArgLysLysThrMetAlaLeuCysPheValIlePheSerPheCys 414
|||||
166 GGGAGGACGAGACCATGGCTTGTGTTTCTTCATCTTCTCCATGTGCAT 215
|||||
414 rLeuLeuLeuPheLeuCysValGlyArg..... 423
::: |||||
216 TATTCCACTCTACGGCTGTGTGCGCGGTAAAGATTATCAGCAGCTGTCA 265
|||||
423 ..... 423
|||||
266 GATTCCAAACTGGAGACCATGACGACATAACTGCCTTTATTTCGCCGNAG 315
|||||
424 AsnValLeuThrLeuLeuPheIleAlaArgAlaPheIleSerGlyG1 440
::: |||||
316 GNCCTCCATGACCGGTGTCATATTATTCATGCCAGACGCTTCATTCAGGAGG 365
|||||
440 yPheGlnAlaAlaTyrValTyrThrProGlu..... 450
|||||
366 ATTTAGGCGCGCTACGTTTACACTCCAGAGTAAAGGACCCAGTTAGAA 415
|||||
451 .....Val 451
|||||
416 TGCAGATGACAGTTTCAAGCCCATCGCTTTGTGCAATCTCTCCAGGTG 465
|||||
452 TyrProThrAlaThrArgAlaLeuGlyLeuGlyThrCysSerGlyMetA1 468
|||||
466 TATCCAACGCGGACGAGGCTTTTGTGCTTGGAAACAGCAGCGAATGGC 515
|||||
468 aArgValGlyAlaLeuLeuThrProPheIleAlaGln..... 480
|||||
516 AAGAGTCGGCGCCTCATCACACCATTTGTGSCACAGGTGAGGACGTCT 565
|||||
480 ..... 480
|||||
566 TTGCTTATTAGCTCAAAACCTCAAAAGATGTGTCGTACACCTGCCGCG 615
|||||
481 .....ValMetLeuGluSerSerValTyrLeuThrLeuAlaVa 493
|||||
616 CTGTGCTTAAAGGTGATGCTGGAGTCTGCTGCTACCTGGCTCTGTTGT 665
|||||
493 lTyrSerGlyCysCysLeuLeuAlaAlaLeuAlaSerCysPheLeuPro1 510
|||||
666 GTACTGCTGCTGCTGCTTCCGCTGCCATCGCTCTCGCGCTGCCCA 715
|||||
510 lGluThrLysGlyGlyLeuGluSerSerHisArgGluTrpGly 526
|||||
716 TCGAGACGAC.GGGTCGGCGCCTGCGAGGTCCANCNAGCACGAGTGGGC 764
|||||
527 GlnGluMetValGlyArgGlyMethisGly.....AlaGlyValThrAr 541
|||||
765 CAGGAAATNGTGGCGCGCCTCGTCCCGCGGTTAGCGAGGANCCTCA 814
|||||
541 gSerAsnSerGlySerGln 547
|||||
815 TTCTCTCTCGGGTCAACA 833
|||||

seq_name: gb_est1:AW345918

```

```
seq_documentation_block:
LOCUS      AW345918      377 bp      mRNA      EST      25-APR-2001
DEFINITION      15969 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION      AW345918
VERSION        AW345918.1      GI:6843628
KEYWORDS      EST.
SOURCE        COW.
ORGANISM      Bos taurus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
              Bovinae; Bovinae; Bos.
REFERENCE      1 (bases 1 to 377)
AUTHORS        Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.F.,
              Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
              ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
              Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
              Keefe,J.W.
TITLE          Sequence evaluation of four pooled-tissue normalized bovine cDNA
              libraries and construction of a gene index for cattle
JOURNAL        Genome Res. 11 (4), 626-630 (2001)
MEDLINE        21180013
COMMENT        Contact: Smith TPL
              USDA, ARS, US Meat Animal Research Center
              PO Box 166, Clay Center, NE 68933-0166, USA
              Tel: 402 762 4366
              Fax: 402 762 4390
              Email: smithemail.marc.usda.gov
              Single pass sequencing. Bases called and trimmed with phred
              v0.980904.e. Vector identified by cross_match with the -minscore 20
              and -minmatch 12 options.
              PCR Primers
              FORWARD: AGGAACAGCTATGACCAT
              BACKWARD: GTTTCCTCCAGTCACGACG
              Plate: 8      row: B      column: 9
              Seq primer: ATTAGGTGACACTATAG.
FEATURES        Location/Qualifiers
source          1..377
                /organism="Bos taurus"
                /db_xref="taxon:9913"
                /clone_lib="MARC 4BOV"
                /tissue_type="pooled"
                /lab_host="DH10B"
                /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                Library made from pooled tissue from day 20 and day 40
                embryos."
BASE COUNT      62 a      112 c      130 g      73 t
ORIGIN
alignment_scores:
Quality: 538.00      Length: 120
Ratio: 4.719      Gaps: 2
Percent Similarity: 95.000      Percent Identity: 91.667
alignment_block:
US-09-911-667A-2 x AW345918
Align seg 1/1 to: AW345918 from: 1 to: 377
431 PheIleAlaArgAlaPheIleSerGlyGlyPheGlnAlaAlaTyrValTy 447
|||||
3 TTCATTGGGGAGCATGGATTCTGGAGGCTTCCAGCAGCCCTACGTTTA 52
447 rThrProGluValTyrProThrAlaThrArgAlaLeuGlyLeuGlyThrc 464
|||||
53 CACACCTGAGGCTTACCCACGCGACCCGAGCGCTGGCGTGGCACCT 102
464 ysSerGlyMetAlaArgValGlyAlaLeuIleThrProPheIleAlaGln 480
|||||
103 GCAGCGCATGGCGAGGTGGGGCCCTCATCACCCCGTTTCATGCTCAG 152
481 ValMet.LeuGlu.SerSerValTyrLeuThrLeuAlaValTyrSerGly 496
||||| :||| ||||||||||||||||||| |||||||
```

```
153 GTGATCGGTGGAGATCATCGGTGTACCTGACGCTGCCGTTACAGTGGC 202
497 CysCysLeuLeuAlaAlaLeuAlaSerCysPheLeuProIleGluThrIly 513
|||||
203 TGTGTGCTCTCTGCTGCCCTGGCCTCCTGCTTTTTCCTTATCGAGACCAA 252
513 sGlyGlyGlyLeuGlnGluSerSerHisArgGluTrpGlyGlnGluMetV 530
|||||
253 AGCCCGCGGACTGCAGGAGTCCAGCCACCGGAATGGGGGCGAGAGATGG 302
530 aiGlyArgGlyMetHisGlyAlaGlyValThrArgSerAsnSerGlySer 546
|||||
303 TTGGCCGAGGGGCACACGCGACAGGTGTGCGCCAGGTGCGAACTCGGGCTCT 352
547 GlnGlu 548
|||||
353 CAGGAG 358
```